

10/081935

(FILE 'HCAPLUS' ENTERED AT 09:29:22 ON 04 DEC 2003)

L29 673 SEA FILE=HCAPLUS ABB=ON PLU=ON COCCIDIOID?
L30 242 SEA FILE=HCAPLUS ABB=ON PLU=ON L29 AND (PROTEIN OR
POLYPEPTIDE OR PROTEIN OR PEPTIDE)
L31 44 SEA FILE=HCAPLUS ABB=ON PLU=ON L30 AND VECTOR
L32 14 SEA FILE=HCAPLUS ABB=ON PLU=ON L31 AND RECOMBIN?

L32 ANSWER 1 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2003:846781 HCAPLUS
TITLE: Cholera vaccine candidate 638: intranasal
immunogenicity and expression of a foreign
antigen from the pulmonary pathogen
Coccidioides immitis
AUTHOR(S): Silva, Anisia J.; Mohan, Archana; Benitez, Jorge
A.
CORPORATE SOURCE: Biochemistry and Immunology, Department of
Microbiology, Morehouse School of Medicine, 720
Westview Dr., SW, Atlanta, GA, 30310-1495, USA
SOURCE: Vaccine (2003), 21(32), 4715-4721
CODEN: VACCDE; ISSN: 0264-410X
PUBLISHER: Elsevier Science Ltd.
DOCUMENT TYPE: Journal
LANGUAGE: English
AB *Vibrio cholerae* strain 638 is a live genetically attenuated
candidate cholera vaccine in which the CTX Φ prophage encoding
cholera toxin has been deleted and hapA, encoding an extracellular
Zn-dependent metalloprotease, was insertionally inactivated. Strain
638 was highly immunogenic when inoculated to adult Swiss mice by
the intranasal route as judged by the induction of a strong serum
vibriocidal antibody response. A side-by-side comparison of strain
638 with its isogenic hapA⁺ precursor (strain 81) in the above model
indicated that inactivation of hapA does not affect immunogenicity.
The spherule-associated antigen 2/proline-rich antigen (Ag2/PRA) of
Coccidioides immitis has been shown to protect mice against
coccidioidomycosis to an extent dependent on the modes of
antigen presentation and challenge with *C. immitis* arthrospores. In
this work, we demonstrate the use of a live genetically attenuated
V. cholerae strain to deliver Ag2/PRA. Ag2/PRA was expressed in 638
as a fusion **protein** with the *Escherichia coli* heat labile
toxin B subunit leader **peptide** using the strong Tac
promoter. The **recombinant** Ag2/PRA was efficiently
expressed, processed and secreted to the periplasmic space.
Intranasal immunizations of adult mice with strain 638 expressing
Ag2/PRA induced serum vibriocidal antibody response to the
vector strain and serum total IgG response to Ag2/PRA.
Strain 638 expressing PRA could be recovered from trachea and lung
up to 20 h after immunization but was effectively cleared 72 h
post-inoculation.

L32 ANSWER 2 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN
ACCESSION NUMBER: 2003:454501 HCAPLUS
DOCUMENT NUMBER: 139:35072
TITLE: **Vectors** comprising nucleotide
sequences for target immunogen, PI31, CIIIA and
antisense HERNA mRNA, and their uses including
use as vaccines
INVENTOR(S): McCreavy, David Thomas; Fraser, William Duncan;
Gallagher, James Anthony

10/081935

PATENT ASSIGNEE(S): University of Liverpool, UK
SOURCE: PCT Int. Appl., 52 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003048371	A2	20030612	WO 2002-GB5512	20021206
WO 2003048371	A3	20030912		

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: GB 2001-29338 A 20011207
GB 2002-23829 A 20021012

AB The invention provides **vectors** (such as viral **vectors**, plasmid **vectors** or phagemids) comprising:
(a) a heterologous nucleotide sequence encoding an antigenic **polypeptide** from a pathogen (such as viral, bacterial, parasitic or fungal); (b) a nucleotide sequence encoding a protease inhibitor (such as human PI31); a nucleotide sequence for a constitutive, regulatable, and/or cell/tissue-specific promoter; and (d) a nucleotide sequence encoding an inhibitory RNA mols., specifically an antisense human HERNA oligonucleotides. The invention also provides **vectors** comprising a nucleotide sequence encoding CIITA, a **polypeptide** that stimulates the expression of MHC class II genes. The invention further provides the use of said **vectors** as vaccines in production of an immune response (humoral) to said antigens in an animal, such as human, wherein said vaccination may be against a viral, fungal, bacterial or parasitic disorder. Still further, the invention relates: (a) using said **vectors** in production of antibodies, wherein said antibodies may be of therapeutic and/or of diagnostic use; (b) that said **vectors** may be adapted for expression of humanized or chimeric antibodies; and (c) that said **vectors** may be used to used to immunize animals for production of hybridomas expressing a monoclonal antibody against antigen of interest. Finally, the invention provides the cDNA sequences of mouse CIITA, and human PI31, and partial cDNA sequence of human HERNA helicase. The invention related that the use of said **vectors** containing said sequences can be used to enhance secretion of translated immunogen, and enhance DNA vaccination bias away from an MHC class I event towards MHC class II event. In the examples, the invention presented the construction of two **vectors**, pcDNAFinal and pcDNA6TR-IRES-CIITA, wherein pcDNAFinal contains nucleotide sequences encoding immunogen parathyroid hormone-related **protein** (PTHrP), antisense HERNA mRNA, PI31 and CD4+ T-cell epitope from lymphocytic choriomeningitis virus and wherein

10/081935

pcDNA6TR-IRES-CIITA encodes CIITA. Specifically, the invention related that: (a) antisense HERNA RNA can increase the transcriptional efficiency of **vectors** resulting in greater levels of transgene expression; (b) PI3l can inhibit proteasome digestion of **recombinant** antigen making it more assessable to MHC class II antigens; (c) inclusion of CD4+ T-cell epitope ensured that degraded immunogen-MHC class II complex bound to CD4+ T cells; and (d) inclusion of CIITA **protein** allowed for over-expression of MHC class II antigens.

L32 ANSWER 3 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2002:637813 HCAPLUS

DOCUMENT NUMBER: 137:180844

TITLE: Yeast Bax-responsive genes for drug target identification in yeast and fungi

INVENTOR(S): Contreras, Roland Henri; Eberhardt, Ines; Luyten, Walter Herman Maria Louis; Reekmans, Rieka Josephina

PATENT ASSIGNEE(S): Janssen Pharmaceutica N.V., Belg.

SOURCE: PCT Int. Appl., 344 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2002064766	A2	20020822	WO 2001-EP15398	20011221
WO 2002064766	A3	20030626		
W:	AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM			
RW:	GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG			

EP 1346044 A2 20030924 EP 2001-997983 20011221

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR

PRIORITY APPLN. INFO.:

EP 2000-870318 A 20001222

EP 2001-870002 A 20010104

EP 2001-870003 A 20010109

WO 2001-EP15398 W 20011221

AB The cDNAs for Saccharomyces cerevisiae genes responding to BAX gene expression as well as the **proteins** encoded by these cDNAs are disclosed. Addnl., Candida albicans and human homologs of the S. cerevisiae genes/**proteins** are provided. The invention describes the use of nucleic acids and **proteins** which are involved in apoptosis in yeast or fungi for the preparation of medicines for treating diseases associated with yeast or fungi or for the treatment of proliferative disorders or for preventing apoptosis in certain diseases. Methods are provided to identify compds. which selectively modulate the expression or functionality of said

10/081935

proteins in the same or a parallel pathway. Also provided are compds. as well as pharmaceutical compns., medicines and vaccines. The invention also comprises new nucleic acid sequences, probes and primers derived thereof, expression **vectors** and host cells transformed with said **vectors**, **polypeptides** and antibodies raised against said **polypeptides**.

L32 ANSWER 4 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2002:489694 HCAPLUS

DOCUMENT NUMBER: 137:92359

TITLE: Role of signal sequence in vaccine-induced protection against experimental **coccidioidomycosis**

AUTHOR(S): Jiang, Chengyong; Magee, D. Mitchell; Ivey, F. Douglas; Cox, Rebecca A.

CORPORATE SOURCE: Department of Microbiology, University of Texas Health Science Center at San Antonio, San Antonio, TX, 78229, USA

SOURCE: Infection and Immunity (2002), 70(7), 3539-3545
CODEN: INFIBR; ISSN: 0019-9567

PUBLISHER: American Society for Microbiology

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The vaccine efficacy of the gene sequence encoding the signal **peptide** of the antigen known as antigen 2 or proline-rich antigen (Ag2/PRA), an immunodominant antigen present in the cell wall of the fungal pathogen **Coccidioides immitis**, was investigated in a murine model of **coccidioidomycosis**. Expression plasmids for Ag2/PRA(1-18) DNA (signal sequence), Ag2/PRA(19-194) DNA (lacking the signal sequence), and Ag2/PRA(1-194) DNA (full length) were inserted in the pVR1012 **vector**, and the constructs were used to vaccinate the highly susceptible BALB/c mouse strain. Immunization with the signal gene sequence significantly reduced the fungal burden in the lungs and spleens of mice 12 days after i.p. challenge with a LD of 2500 C. immitis arthroconidia, to a level comparable to the protection induced in mice immunized with the full-length Ag2/PRA(1-194) DNA. The Ag2/PRA(19-194) gene protected mice but to a significantly lower level than the signal sequence or the full-length Ag2 gene. The immunizing capacity of Ag2/PRA(1-18) was not attributable to a nonspecific immunostimulatory effect of DNA, as evidenced by the fact that mice immunized with a frameshift mutation of Ag2/PRA(1-18) were not protected against challenge. Furthermore, a synthetic **peptide** corresponding to the translated sequence of Ag2/PRA(1-18) DNA protected mice, albeit at a lower level than the Ag2/PRA(1-18) DNA vaccine. The protection induced with the signal gene vaccine correlated with the production of gamma interferon when splenocytes from Ag2/PRA(1-18)-immunized mice were stimulated with **recombinant** full-length Ag2 and was not associated with the production of anti-**Coccidioides** IgG antibody. This is the first study to establish that a signal **peptide** sequence alone, administered as a gene vaccine or synthetic **peptide**, can induce protective immunity against a microbial pathogen.

REFERENCE COUNT: 36 THERE ARE 36 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

10/081935

L32 ANSWER 5 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2002:366877 HCAPLUS

DOCUMENT NUMBER: 137:137416

TITLE: Complementation of the ODC-Escherichia coli (EWH319) null mutant by expression of the **Coccidioides immitis** ornithine decarboxylase (CiODC) gene

AUTHOR(S): Guevara-Olvera, L.; Guevara-Gonzalez, R. G.; Munoz-Sanchez, C. I.; Gonzalez-Chavira, M. M.

CORPORATE SOURCE: Inst. Tecnológico de Celaya, Depto. Ingeniería Bioquímica, Guanajuato, 38010, Mex.

SOURCE: Informacion Tecnologica (2002), 13(2), 177-181
CODEN: ITECFG; ISSN: 0716-8756

PUBLISHER: Centro de Informacion Tecnologica

DOCUMENT TYPE: Journal

LANGUAGE: Spanish

AB The aim of this study was to determine the function of the ornithine decarboxylase (CiODC) gene from **Coccidioides immitis**, a human respiratory fungal pathogen. A fragment of complementary DNA (cDNA) coding for CiODC **protein** was cloned into the pET-28b **vector**. The pET-CiODC **recombinant** plasmid was used to transform the EWH319 strain of Escherichia coli, which lacks the ODC gene and all others which synthesize polyamines and shows reduced growth due to this deficiency. Introduction of the transformant plasmid restored the bacterial growth in minimal M9 medium. The complemented strain EWH319.pET-CiODC showed ODC activity which was inhibited by 1,4-diamino-2-butanone (DAB) a competitive inhibitor which is specific for eukaryote ODCs. Based on the present results it was concluded that expression of the CiODC gene from **Coccidioides immitis** conferred a fundamental function for the growth of Escherichia coli.

REFERENCE COUNT: 20 THERE ARE 20 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L32 ANSWER 6 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2002:107512 HCAPLUS

DOCUMENT NUMBER: 136:162279

TITLE: Production of human monoclonal antibodies in human B-lymphocyte hybridomas expressing an ectopic telomerase gene

INVENTOR(S): Dessain, Scott K.; Goldsby, Richard A.

PATENT ASSIGNEE(S): Whitehead Institute for Biomedical Research, USA

SOURCE: PCT Int. Appl., 74 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2002010352	A2	20020207	WO 2001-US24591	20010801
WO 2002010352	A3	20030227		

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,

10/081935

NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR,
TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ,
MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH,
CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE,
TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN,
TD, TG

US 2002045219 A1 20020418 US 2001-759984 20010112
PRIORITY APPLN. INFO.: US 2000-222473P P 20000802
US 2001-759984 A1 20010112

AB The present invention relates to a method of making human monoclonal antibodies through the use of novel hybrid cells. In the invention, hybrid cells are created by combining three elements: a fusion partner cell, a fusion cell (in particular a human B-lymphocyte), and an ectopic telomerase gene. Mammalian cell lines that ectopically express telomerase and methods of using such cell lines in producing novel hybrid cells (hybridomas) that produce human monoclonal antibodies; human monoclonal antibodies produced by such novel hybridomas and DNA constructs useful for producing mammalian cell lines that ectopically express telomerase are described. The expression of an ectopic telomerase gene in hybrid cells formed from primary human B-lymphocytes and fusion partner cells (other human non-B lineage cells) improves their growth rate, level of Ig expression, stability of Ig expression, and the ability to be cloned by limiting dilution. A murine myeloma cell line that ectopically expressed human telomerase was created, as well as murine/human cell hybrids.

L32 ANSWER 7 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2001:935789 HCAPLUS

DOCUMENT NUMBER: 136:65197

TITLE: Sequences of antigenic **polypeptides** of
staphylococcus aureus and their uses in against
bacterial infection

INVENTOR(S): Foster, Simon; McDowell, Philip; Brummell,
Kirsty; Clarke, Simon

PATENT ASSIGNEE(S): University of Sheffield, UK; Biosynexus Inc.

SOURCE: PCT Int. Appl., 85 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001098499	A1	20011227	WO 2001-GB2685	20010620
W:	AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM			
RW:	GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG			

10/081935

EP 1292681 A1 20030319 EP 2001-940746 20010620
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC,
PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR
BR 2001011823 A 20030610 BR 2001-11823 20010620
NO 2002005838 A 20030218 NO 2002-5838 20021205
US 2003186275 A1 20031002 US 2003-311879 20030318
PRIORITY APPLN. INFO.: GB 2000-14907 A 20000620
WO 2001-GB2685 W 20010620

AB The invention discloses methods for the identification of antigenic **proteins** expressed by pathogenic microbes, vaccines comprising the **proteins**, **recombinant** methods to manufacture the **proteins** and therapeutic antibodies directed to the **proteins**. In particular, the invention discloses amino acid sequences of staphylococcus aureus antigenic **proteins**, the DNA sequences encoding **polypeptides** and genomic DNA library of staphylococcus aureus. The invention also provides expression **vectors** encoding antigenic **peptides**, methods for the production of the **proteins**, antibodies to the **proteins** as well as methods of preparing the antibodies. The invention further provides vaccine comprising the antigenic **proteins**, pharmaceutical carrier, and adjuvant as well as methods of immunizing animals or humans.

REFERENCE COUNT: 2 THERE ARE 2 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L32 ANSWER 8 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2001:309883 HCAPLUS

DOCUMENT NUMBER: 135:91197

TITLE: **Recombinant** urease and urease DNA of **Coccidioides immitis** elicit an immunoprotective response against **coccidioidomycosis** in mice

AUTHOR(S): Li, Kun; Yu, Jieh-Juen; Hung, Chiung-Yu; Lehmann, Paul F.; Cole, Garry T.

CORPORATE SOURCE: Department of Microbiology and Immunology, Medical College of Ohio, Toledo, OH, 43614-5806, USA

SOURCE: Infection and Immunity (2001), 69(5), 2878-2887
CODEN: INFIBR; ISSN: 0019-9567

PUBLISHER: American Society for Microbiology

DOCUMENT TYPE: Journal

LANGUAGE: English

AB **Coccidioides immitis** antigens which stimulate a T helper cell 1 (Th1) pathway of host immune response are considered to be essential components of a vaccine against **coccidioidomycosis**. **Recombinant** urease (rURE) and **recombinant** heat shock **protein** 60 (rHSP60) of *C. immitis* were expressed in *Escherichia coli* and tested as vaccine candidates in BALB/c mice. A synthetic oligodeoxynucleotide which contained unmethylated CpG dinucleotides and was previously shown to enhance a murine Th1 response was used as an immunoadjuvant. T cells isolated from the spleens and lymph nodes of the rURE- and rHSP60-immune mice showed in vitro proliferative responses to the resp. **recombinant protein**, but only those T lymphocytes from rURE-immunized mice revealed markedly elevated levels of expression of selected Th1-type cytokine genes. BALB/c mice immunized s.c. with rURE and subsequently challenged by the i.p.

10/081935

route with a lethal inoculum of *C. immitis* arthroconidia demonstrated a significant reduction in the level of *C. immitis* infection compared to control animals. RHSP60 was much less effective as a protective antigen. Evaluation of cytokine gene expression in lung tissue and levels of **recombinant** urease-specific Igs (IgG1 vs. IgG2a) in murine sera at 12 days after challenge provided addnl. evidence that immunization with rURE stimulated a Th1 response to the pathogen. Urease was further evaluated by expression of the URE gene in a mammalian plasmid **vector** (pSecTag2A.URE) which was used to immunize mice by the intradermal route. In this case, 82% of the **vector** construct-immunized animals survived more than 40 days after i.p. infection, compared to only 10% of the mice immunized with the **vector** alone. In addition, 87% of the pSecTag2A.URE-immunized survivors had sterile lungs and spleens. These data support the need for further evaluation of the *C. immitis* urease as a candidate vaccine against **coccidioidomycosis**.

REFERENCE COUNT: 44 THERE ARE 44 CITED REFERENCES AVAILABLE
FOR THIS RECORD. ALL CITATIONS AVAILABLE
IN THE RE FORMAT

L32 ANSWER 9 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 1999:83280 HCAPLUS

DOCUMENT NUMBER: 130:266091

TITLE: Genetic vaccination against **Coccidioides**
immitis: comparison of vaccine efficacy of
recombinant antigen 2 and antigen 2 cDNA

AUTHOR(S): Jiang, Chengyong; Magee, D. Mitchell; Quitugua,
Teresa N.; Cox, Rebecca A.

CORPORATE SOURCE: Department of Clinical Investigation, Texas
Center for Infectious Disease, San Antonio, TX,
78223, USA

SOURCE: Infection and Immunity (1999), 67(2), 630-635
CODEN: INFIBR; ISSN: 0019-9567

PUBLISHER: American Society for Microbiology

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Previous studies from our laboratory established that C-ASWS, an alkali-soluble, water-soluble extract from cell walls of **Coccidioides immitis**, protects mice against lethal challenge with this fungus. The C-ASWS extract contains a glycosylated **protein**, designated antigen 2 (Ag2), and a polysaccharide antigen. We recently cloned Ag2 cDNA and showed that the **recombinant** fusion **protein** elicited strong delayed-type hypersensitivity responses in immunized mice. This investigation was undertaken to determine if the **recombinant** Ag2 **protein**, expressed as an Ag2-glutathione S-transferase (GST) fusion **protein**, or Ag2 cDNA would protect mice against lethal challenge with *C. immitis*. The **recombinant** Ag2-GST **protein** protected BALB/c mice against i.p. challenge with 250 arthroconidia, as assessed by a decrease in fungal CFU in tissues. The Ag2-GST-immunized mice did not show, however, an increased survival during a 30-day period postinfection. By contrast, immunization of mice with Ag2 cDNA ligated into the pVR1012 plasmid engendered protection against i.p. challenge with 2,500 arthroconidia and against pulmonary challenge with 50 arthroconidia. Vaccine efficacy paralleled the development of delayed-type hypersensitivity responses to *C. immitis* antigen.

10/081935

Whereas mice vaccinated with the **recombinant** Ag2-GST **protein** did not mount footpad hypersensitivity to C-ASWS or the **recombinant** Ag2-GST **protein**, mice vaccinated with the pVR1012-Ag2 construct mounted a strong footpad hypersensitivity and their spleen cells secreted gamma interferon upon in vitro stimulation with the Ag2-containing C-ASWS extract This is the first investigation to show that genetic immunization can protect against lethal challenge with *C. immitis*.

REFERENCE COUNT: 50 THERE ARE 50 CITED REFERENCES AVAILABLE
FOR THIS RECORD. ALL CITATIONS AVAILABLE
IN THE RE FORMAT

L32 ANSWER 10 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 1999:18776 HCAPLUS

DOCUMENT NUMBER: 130:236203

TITLE: Detecting serum antibodies to a purified
recombinant proline-rich antigen of
Coccidioides immitis in patients with
coccidioidomycosis

AUTHOR(S): Orsborn, Kris I.; Galgiani, John N.

CORPORATE SOURCE: Research and Medical Services, Veterans Affairs
Medical Center, University of Arizona, Tucson,
AZ, USA

SOURCE: Clinical Infectious Diseases (1998), 27(6),
1475-1478

CODEN: CIDIEL; ISSN: 1058-4838

PUBLISHER: University of Chicago Press

DOCUMENT TYPE: Journal

LANGUAGE: English

AB In previous work, antibodies in serum samples from patients with **coccidioidomycosis** were found to react with a proline-rich antigen (PRA) isolated from spherules of **Coccidioides immitis**, and the gene encoding this antigen was cloned. We expressed and purified **recombinant** PRA (rPRA) by removing the majority of amino acids contributed by the **vector** from the fusion **protein**. Purified rPRA reacted with serum IgG antibodies in 37 of 42 patients with culture-proven progressive pulmonary or extrapulmonary **coccidioidal** disease; specific antibodies in dilns. ranging from 1:40 to 1:102,400 were demonstrated (sensitivity, 88%). In contrast, for >95% of patients without **coccidioidomycosis** reactivity of <1:40 was demonstrated (specificity, 97%). Of 18 patients with primary self-limited **coccidioidomycosis**, none had detectable antibodies in serum samples collected up to 141 days after illness began. The association of antibodies to rPRA with progressive infection may have prognostic value.

REFERENCE COUNT: 22 THERE ARE 22 CITED REFERENCES AVAILABLE
FOR THIS RECORD. ALL CITATIONS AVAILABLE
IN THE RE FORMAT

L32 ANSWER 11 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 1996:391996 HCAPLUS

DOCUMENT NUMBER: 125:55750

TITLE: Molecular cloning and characterization of
Coccidioides immitis antigen 2 cDNA

AUTHOR(S): Zhu, Yufan; Yang, Chunmu; Magee, D. Mitchell;
Cox, Rebecca A.

CORPORATE SOURCE: Dep. Clinical Investigation, Texas Center

10/081935

SOURCE: Infectious Disease, San Antonio, TX, 78223, USA
Infection and Immunity (1996), 64(7), 2695-2699
CODEN: INFIBR; ISSN: 0019-9567
PUBLISHER: American Society for Microbiology
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Previous expts. have provided evidence that **Coccidioides** immitis antigen 2 (Ag2) is a major T-cell-reactive component of mycelia and spherule cell walls. Here the authors report the identification and cloning of the cDNA that encodes Ag2 from a lambda ZAP cDNA expression library constructed from spherule-derived RNA. DNA sequence anal. established that the 1,255-bp clone contains a 174-bp 5' untranslated region, a 582-bp open reading frame which encodes for a **protein** consisting of 194 amino acids, and a 375-bp 3' untranslated region, including a poly(A) tail. The **recombinant** Ag2 **protein** has a predicted mol. mass of 19.5 kDa and contains an 18-amino-acid N terminus which has been tentatively identified as a signal **peptide**. The Ag2 cDNA was ligated into the pGEX-4T-3 **vector** and expressed in Escherichia coli TG-1 cells as a glutathione S-transferase fusion **protein**. The **recombinant** fusion **protein** showed reactivity with sera from patients with **coccidioidomycosis** and elicited delayed-type footpad hypersensitivity responses in **Coccidioides**-immune mice. These results suggest that the Ag2 cDNA can be used for the large-scale production of this immunol. important **protein**.

L32 ANSWER 12 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 1996:327965 HCAPLUS
DOCUMENT NUMBER: 125:28509
TITLE: Molecular cloning and characterization of the **Coccidioides** immitis complement fixation/chitinase antigen
AUTHOR(S): Yang, Chunmu; Zhu, Yufan; Magee, D. Mitchell; Cox, Rebecca A.
CORPORATE SOURCE: Department Clinical Investigation, Texas Center Infectious disease, San Antonio, TX, 78223, USA
SOURCE: Infection and Immunity (1996), 64(6), 1992-1997
CODEN: INFIBR; ISSN: 0019-9567
PUBLISHER: American Society for Microbiology
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Detection of anti-**Coccidioides** complement-fixing (CF) antibody is a valuable diagnostic and prognostic aid in **coccidioidomycosis**. The CF antibody response is directed against a heat-labile antigen that has chitinase activity, hereafter referred to as the CF/chitinase **protein**. To identify and clone this immunoreactive enzyme, the authors constructed a **Coccidioides** immitis cDNA lambda ZAP expression library from spherule RNA and detected fusion **peptides** expressing CF epitopes by immunoscreening. A cDNA clone consisting of 1623 bp was identified, sequenced, and found to contain a single open reading frame that encodes a **protein** of 47 kDa with 427 amino acids. Deduced amino acid sequence analyses showed that the cloned CF/chitinase cDNA contains a 35-amino-acid region, beginning at Ser-18 and ending at Arg-52, which has 92% homol. with the reported N-terminal amino acid sequence of authentic CF/chitinase

10/081935

protein. The first 17 amino acids in the deduced sequence of the cloned cDNA are not present on the mature CF/chitinase **protein**, suggesting that it may be a signal **peptide**

. Expression of the CF/chitinase cDNA insert by using the pGEX-4T-3 **vector** yields a fusion **peptide** that bears CF-specific epitopes and shows chitinase activity. The CF/chitinase clone will enable large-scale production of the **recombinant** CF antigen for use in immunoassays and facilitate studies on the role of chitinase in the morphogenesis of *C. immitis*.

L32 ANSWER 13 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 1995:832234 HCAPLUS

DOCUMENT NUMBER: 124:2086

TITLE: Molecular and biochemical characterization of a **Coccidioides immitis**-specific antigen

AUTHOR(S): Pan, Shuchong; Cole, Garry T.

CORPORATE SOURCE: Department of Microbiology, Medical College of Ohio, Toledo, OH, 43699-008, USA

SOURCE: Infection and Immunity (1995), 63(10), 3994-4002
CODEN: INFIBR; ISSN: 0019-9567

PUBLISHER: American Society for Microbiology

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Results of earlier investigations have indicated that the saprobic phase of **Coccidioides immitis** produces a heat-stable, 19-kDa antigen with serine proteinase activity which has been suggested to be specific for this pathogenic fungus. In the present study the authors have determined the N-terminal and partial internal amino acid sequences of the purified, 19-kDa antigen, cloned the gene which encodes this **polypeptide**, and confirmed that the secreted proteinase is a **Coccidioides**-specific antigen (CS-Ag). Both the genomic and cDNA sequences are reported and reveal that the *csa* gene which encodes this antigen has no introns. A 543-bp open reading frame encodes a 181-amino-acid-containing **protein** with a predicted mol. mass of 19.8 kDa and an isoelec. point of 8.3. The *csa* gene was localized on chromosome I of three representative *C. immitis* clin. isolates on the basis of Southern hybridizations. Expression of the *csa* gene in *Escherichia coli* using the pET21a plasmid **vector** yielded a **recombinant protein** that was recognized in immunoblot assays by antibody raised to the purified 19-kDa CS-Ag. Secretion of the native antigen is suggested to occur by cleavage of a putative 23-residue signal **peptide**. The native CS-Ag showed a low degree of glycosylation. Anal. of the carbohydrate composition of the CS-Ag revealed xylose, mannose, galactose, and glucose. However, the purified antigen showed no affinity for Con A. A PCR method with specificity and high sensitivity for detection of *C. immitis* genomic DNA, using a pair of synthetic oligonucleotide primers whose sequences were based on that of the *csa* gene, was developed. A 520-bp product was amplified only when *C. immitis* genomic DNA was used as the template. The lower limits of DNA detection using this PCR method were 1 pg of *C. immitis* genomic DNA by ethidium bromide staining and 100 fg after Southern hybridization. The *csa* gene-based PCR method for detection of *C. immitis* DNA is useful for culture identification and may have clin. applications for the diagnosis of **coccidioidal** infections.

L32 ANSWER 14 OF 14 HCAPLUS COPYRIGHT 2003 ACS on STN

10/081935

ACCESSION NUMBER: 1994:267638 HCAPLUS
DOCUMENT NUMBER: 120:267638
TITLE: Identification of antigens of
Coccidioides immitis which stimulated
immune T lymphocytes
AUTHOR(S): Cole, Garry T.; Kirkland, Theo N.
CORPORATE SOURCE: Dep. Bot., Univ. Texas, Austin, TX, 78713, USA
SOURCE: Archives of Medical Research (1993), 24(3),
281-91
CODEN: AEDEER; ISSN: 0188-4409
DOCUMENT TYPE: Journal
LANGUAGE: English
AB T-cell mediated immune response to **coccidioidomycosis** has
been shown to be the principal mechanism of resistance to this
respiratory fungal disease in exptl. animals. In this study, a
Coccidioides immitis antigen-specific murine T-cell line was
used to identify macromols. capable of eliciting an immune mouse
T-cell proliferative response. The murine T-cell line was selected
on the basis of its strong pos. response to a soluble conidial wall
fraction (SCWF), which had previously been shown to be reactive in
humoral and cellular immunoassays. An antigen-specific T-cell line
rather than T-cell clones was used to identify multiple antigens.
The T-cell immunoblot method was employed first to identify
immunoreactive subfractions of the SCWF, and then to identify T-cell
fusion **proteins** (FPs) obtained from a cDNA expression
library constructed in λ gt11. The library was screened with
anti-SCWF. The nucleotide sequence of a 0.2-kb cDNA insert encoding
a FP which elicited vigorous T-cell response was determined. A construct
of this insert was subcloned into the pET expression **vector**
system and 6.5-kDa **recombinant protein** (RP)
expressed in *Escherichia coli* was isolated. The RP and FP were
shown to be homologous on the basis of identify of their amino acid
sequences. Antibody raised in guinea pigs against the RP recognized
a 59-kDa native **protein** of the mycelial culture filtrate
produced by 3 sep. strains of *C. immitis*, and reacted with the cell
wall of arthroconidia as detected by immunofluorescence microscopy.
In this study, a potentially important T-cell stimulating antigen of
C. immitis was identified and partly characterized.

(FILE 'MEDLINE, BIOSIS, EMBASE, WPIDS, CONFSCI, SCISEARCH,
JICST-EPLUS, JAPIO' ENTERED AT 09:34:38 ON 04 DEC 2003)
L33 37 S L32
L34 18 DUP REM L33 (19 DUPLICATES REMOVED)

L34 ANSWER 1 OF 18 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2003-829776 [77] WPIDS
CROSS REFERENCE: 1999-551417 [46]; 2002-606631 [65]
DOC. NO. NON-CPI: N2003-662909
DOC. NO. CPI: C2003-233753
TITLE: Novel chitin-binding, chitinase-inactive
polypeptide comprising a chitin-binding
fragment, useful for treating fungal infections.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): GRAY, P W; TJOELKER, L W
PATENT ASSIGNEE(S): (GRAY-I) GRAY P W; (TJOE-I) TJOELKER L W
COUNTRY COUNT: 1
PATENT INFORMATION:

10/081935

PATENT NO	KIND	DATE	WEEK	LA	PG
US 2003143216	A1	20030731	(200377)*		33

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
US 2003143216	A1 CIP of	US 1998-39198	19980312
	Div ex	US 1999-267574	19990312
		US 2002-161547	20020603

FILING DETAILS:

PATENT NO	KIND	PATENT NO
US 2003143216	A1 CIP of	US 6200951
	Div ex	US 6399571

PRIORITY APPLN. INFO: US 1999-267574 19990312; US 1998-39198
19980312; US 2002-161547 20020603

AN 2003-829776 [77] WPIDS
CR 1999-551417 [46]; 2002-606631 [65]
AB US2003143216 A UPAB: 20031128

NOVELTY - A chitin-binding, chitinase-inactive **polypeptide** (I) comprising a chitin-binding fragment of 54 C-terminal amino acids of a fully defined human chitinase sequence (S1) of 466 amino acids as given in the specification, is new. (I) is chosen from **polypeptides** having sequence of amino acid residues 347-397 to 445 of (S1).

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for:

- (1) a fusion **protein** (II) comprising (I) fused to a heterologous **polypeptide**;
- (2) a composition (III) comprising (I) and a diluent;
- (3) a composition (IV) comprising (I) conjugated to an antifungal agent;
- (4) a composition (V) comprising (I) conjugated to a detectable label;
- (5) a kit (VI) for diagnosing the presence of chitin in a sample comprising the composition;
- (6) a purified, isolated polynucleotide (VII) encoding (I);
- (7) a **vector** (VIII) comprising (VII);
- (8) a host cell (IX) transformed or transfected with (VII) in a manner allowing the expression of (I) encoded by (VII) in (IX);
- (9) a **polypeptide** produced by culturing (IX); and
- (10) a monoclonal antibody (X) that specifically binds to an epitope within the 54C-terminal amino acid of (S1).

ACTIVITY - Fungicide.

Anti-fungal activity of **recombinant** chitinase in vivo in mice was determined as follows. Female Balb/c mice, 6-8 weeks old, were administered 0.5, 5.0, 50 mg/kg **recombinant** human chitinase by intravenous injection in the tail vein. The mice was infected with the fungi, varying doses of test drug was administered to the animals and their survival is measured over time. Specifically, acute systemic candidiasis was achieved in mice by intraperitoneal or intravenous challenge of 10 multiply 106 colony forming units (CFU) candida albicans. The **recombinant**

10/081935

chitinase was administered before or at 1 to 5 hours after challenge, and the number of survivors was determined after five days. In addition, the mice can be sacrificed and fungal load can be determined in specific organs. Alternatively, the mice were challenged with lower doses of fungi, e.g., *Aspergillus* (8-10 multiply 10⁶ CFU) or *Candida* (1 multiply 10⁶ CFU) in which case survival can be measured at more distant time points, e.g. 45 days. The long term fungicidal/fungistatic activity of a chitinase may be evaluated by continuing therapy for a week or more, e.g. 11 days, and following the animals over several weeks, e.g. 18 days to one month. The result showed that effective anti-fungal agent (**recombinant** chitinase) enhanced the long term survival of animals and reduced fungal load in blood and organs.

MECHANISM OF ACTION - Lysis of fungal cell wall and membrane.

USE - (III) or (IV) is useful for treating (M1) fungal infection which involves administering (III) or (IV) to a subject suffering from fungal infection. (M1) further comprises administering a non-chitinase antifungal agent. (V) is useful for detecting the presence of chitin in a sample which involves contacting the sample with (V) and determining the amount of labeled **polypeptide** bound to chitin. (claimed).

(III), (IV) or (V) is useful for treating fungal infections such as candidiasis, aspergillosis, **coccidioidomycosis**, blastomycosis, paracoccidioidomycosis, histoplasmosis, cryptococcosis, chromoblastomycosis, sporotrichosis, mucormycosis, and the dermatophytoses can manifest as acute or chronic disease. (VII) is useful in hybridization assays to detect the capacity of cells to synthesize chitinase and for diagnostic methods useful for identifying a genetic alteration(s) in the chitinase locus that underlies a disease state or states. (X) is useful for detecting or quantifying the presence of chitin-binding domain, and for detecting or quantifying the presence of yeast or fungi, e.g., by adding a chitin-binding domain which binds to the yeast or fungi. (X) is also useful for detecting chitin-binding domain in human blood samples which indicates a disease state involving chitinase, such as Gaucher's disease.

Dwg.0/0

L34 ANSWER 2 OF 18 MEDLINE on STN DUPLICATE 1
ACCESSION NUMBER: 2003510266 IN-PROCESS
DOCUMENT NUMBER: 22949199 PubMed ID: 14585681
TITLE: Cholera vaccine candidate 638: intranasal
immunogenicity and expression of a foreign antigen
from the pulmonary pathogen **Coccidioides**
immitis.
AUTHOR: Silva Anisia J; Mohan Archana; Benitez Jorge A
CORPORATE SOURCE: Department of Microbiology, Biochemistry and
Immunology, Morehouse School of Medicine, 720
Westview Dr., SW 30310-1495, Atlanta, GA, USA.
SOURCE: VACCINE, (2003 Dec 1) 21 (32) 4715-21.
Journal code: 8406899. ISSN: 0264-410X.
PUB. COUNTRY: Netherlands
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: IN-PROCESS; NONINDEXED; Priority Journals
ENTRY DATE: Entered STN: 20031031
Last Updated on STN: 20031031
AB *Vibrio cholerae* strain 638 is a live genetically attenuated

10/081935

candidate cholera vaccine in which the CTXPhi prophage encoding cholera toxin has been deleted and hapA, encoding an extracellular Zn-dependent metalloprotease, was insertionally inactivated. Strain 638 was highly immunogenic when inoculated to adult Swiss mice by the intranasal route as judged by the induction of a strong serum vibriocidal antibody response. A side-by-side comparison of strain 638 with its isogenic hapA(+) precursor (strain 81) in the above model indicated that inactivation of hapA does not affect immunogenicity. The spherule-associated antigen 2/proline-rich antigen (Ag2/PRA) of *Coccidioides immitis* has been shown to protect mice against *coccidioidomycosis* to an extent dependent on the modes of antigen presentation and challenge with *C. immitis* arthrospores. In this work, we demonstrate the use of a live genetically attenuated *V. cholerae* strain to deliver Ag2/PRA. Ag2/PRA was expressed in 638 as a fusion **protein** with the *Escherichia coli* heat labile toxin B subunit leader **peptide** using the strong Tac promoter. The **recombinant** Ag2/PRA was efficiently expressed, processed and secreted to the periplasmic space. Intranasal immunizations of adult mice with strain 638 expressing Ag2/PRA induced serum vibriocidal antibody response to the **vector** strain and serum total IgG response to Ag2/PRA. Strain 638 expressing PRA could be recovered from trachea and lung up to 20h after immunization but was effectively cleared 72h post-inoculation.

L34 ANSWER 3 OF 18 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2003-120579 [11] WPIDS
DOC. NO. NON-CPI: N2003-096057
DOC. NO. CPI: C2003-031163
TITLE: Identifying biologically active agents comprises cloning transfected cells into a cell array, exposing the array to an agent to be tested, and detecting signals generated by a reporter molecule as a result of exposure to the agent.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): ANDREWS, P; DRAPER, J; WALSH, J
PATENT ASSIGNEE(S): (AXOR-N) AXORDIA LTD
COUNTRY COUNT: 100
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG

WO 2002090992	A2	20021114	(200311)*	EN	90
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC					
MW MZ NL OA PT SD SE SL SZ TR TZ UG ZM ZW					
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ					
DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP					
KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ					
NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ					
UA UG US UZ VN YU ZA ZM ZW					

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE

WO 2002090992	A2	WO 2002-GB1946	20020429

10/081935

PRIORITY APPLN. INFO: GB 2001-11004 20010504

AN 2003-120579 [11] WPIDS

AB WO 200290992 A UPAB: 20030214

NOVELTY - Screening (M1) for identifying biologically active agents, comprises:

(i) providing a population of cells stably transfected with a nucleic acid encoding a reporter molecule;

(ii) cloning the transfected cells into a cell array;

(iii) exposing the array to at least one agent to be tested;

and

(iv) detecting a signal generated by the reporter molecule as a result of exposure to the agent.

DETAILED DESCRIPTION - Screening (M1) for identifying biologically active agents, comprises:

(i) providing a population of cells which have been stably transfected with a nucleic acid molecule encoding a reporter molecule;

(ii) cloning the transfected cells into a cell array;

(iii) exposing the array to at least one agent to be tested;

and

(iv) detecting a signal generated by the reporter molecule as a result of exposure to the agent.

INDEPENDENT CLAIMS are also included for the following:

(1) An agent identified by M1;

(2) A cell or a cell array obtained by M1;

(3) Screening (M2) for the isolation of a gene, comprising:

(a) steps (i)-(iv) of M1;

(b) extracting nucleic acid from a cell sample comprising the cell array; and

(c) determining the sequence of at least part of the genomic region into which the nucleic acid encoding the reporter molecule has integrated;

(4) Comparing the biological activity of a reference agent with at least one other agent, comprising:

(a) steps (i) and (ii) of the above method;

(b) preparing a duplicate array;

(c) step (iii) of the above method;

(d) exposing the duplicate array to a reference agent; and

(e) detecting a signal generated by the reporter molecule as a result of exposure to the agent and to the reference agent; and

(5) A **vector** comprising a reporter molecule, a splice acceptor site and an internal ribosome entry site, where the splice acceptor and the internal ribosome entry site are operably linked to facilitate expression of the reporter molecule.

USE - The method is useful in identifying biologically active agents and the genes through which the agents act, in screening potential drugs for their ability to activate certain drug targets in a high-throughput assay, in identifying relationships between signaling pathways and specific signals that could be useful in eventually directing the differentiation of embryonic stem cells, and in toxicology assays by testing for unwanted activation or inhibition of specific signaling pathways. The **vector** is useful in carrying out the above methods (claimed).

Dwg.0/99

L34 ANSWER 4 OF 18 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN

ACCESSION NUMBER: 2002-566694 [60] WPIDS

DOC. NO. CPI: C2002-160652

10/081935

TITLE: Constructing strains for identifying gene products as effective targets for therapeutic intervention, by inactivating in the strain one allele of a gene and placing other allele of the gene under conditional expression.

DERWENT CLASS: B04 C06 D16

INVENTOR(S): BOONE, C; BUSSEY, H; JIANG, B; OHLSEN, K L; ROEMER, T

PATENT ASSIGNEE(S): (ELIT-N) ELITRA PHARM INC

COUNTRY COUNT: 101

PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2002053728	A2	20020711	(200260)*	EN	167
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZM ZW					
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TN TR TT TZ UA UG US UZ VN YU ZA ZW					
US 2003180953	A1	20030925	(200364)		
EP 1348027	A2	20031001	(200365)	EN	
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI TR					

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2002053728	A2	WO 2001-US49486	20011226
US 2003180953	A1	Provisional	US 2000-259128P 20001229
		CIP of	US 2001-792024 20010220
		Provisional	US 2001-314050P 20010822
			US 2001-32585 20011220
EP 1348027	A2	EP 2001-991419	20011226
		WO 2001-US49486	20011226

FILING DETAILS:

PATENT NO	KIND	PATENT NO
EP 1348027	A2 Based on	WO 2002053728

PRIORITY APPLN. INFO: US 2001-314050P 20010822; US 2000-259128P 20001229; US 2001-792024 20010220; US 2001-32585 20011220

AN 2002-566694 [60] WPIDS

AB WO 200253728 A UPAB: 20020919

NOVELTY - Constructing (M1) a strain of diploid fungal cells in which both alleles of a gene are modified, comprising modifying one allele by insertion or replacement by a cassette having an expressible selectable marker, and modifying other allele by **recombination**, of a promoter replacement fragment with a heterologous promoter, so that expression of the second allele is regulated by the promoter, si new.

DETAILED DESCRIPTION - Constructing (M1) a strain of diploid

fungal cells in which both alleles of a gene are modified, comprising constructing a strain of diploid fungal cells in which both alleles of a gene are modified, comprises modifying one allele in diploid fungal cells by **recombination** using a gene disruption cassette comprising a first nucleotide sequence encoding an expressible selectable marker, to provide heterozygous diploid fungal cells in which the first allele is inactivated, and modifying the second allele by **recombination** using a promoter replacement fragment comprising a second nucleotide sequence encoding a heterologous promoter, so that expression of the second allele is regulated by the heterologous promoter, where the gene encodes a **polypeptide** required for proliferation of *Candida albicans*, where the **polypeptide** is chosen from the amino acid sequence (S1) of 932 essential genes listed in the specification. INDEPENDENT CLAIMS are also included for the following:

(1) assembling (M2) a collection of diploid fungal cells each comprising modified alleles of a different gene, by modifying both alleles of the gene by (M1) and repeating the steps several times, where a different gene is modified with each repetition;

(2) a strain (I) of diploid fungal cells comprising modified alleles of a gene, where the first allele is inactivated by a gene disruption cassette comprising a nucleotide sequence encoding an expressible selectable marker, and the expression of the second allele is regulated by a heterologous promoter that is operably linked to the coding region of the second allele;

(3) a collection of diploid fungal strains comprising (I);

(4) a nucleic acid molecule microarray comprising several nucleic acid molecules, each comprising a nucleotide sequence that is hybridizable to a target nucleotide sequence chosen from 932 nucleotide sequences (S2) of the open reading frame (ORF) of essential genes, gene names listed in the specification, no sequences are given;

(5) a purified or isolated nucleic acid molecule (II) comprising a nucleotide sequence encoding a gene product required for proliferation of *C. albicans*, and comprising (S1);

(6) a nucleic acid molecule (III) comprising at least 10-100 consecutive nucleotides of (S2);

(7) a purified or isolated nucleic acid molecule (IV) obtained from an organism other than *C. albicans* or *Saccharomyces cerevisiae* comprising at least 30 % identity to a (S2), as determined using BLASTN version 2.0 with the default parameters;

(8) a **vector** containing the promoter operably linked to (II), (III) or (IV);

(9) a host cell containing the **vector** of (9);

(10) a purified or isolated **polypeptide** comprising a sequence chosen from 61 sequences not given in the specification;

(11) a purified or isolated **polypeptide** (V) obtained from an organism other than *C. albicans* or *S. cerevisiae* comprising at least 30 % similarity to (S1), as determined using FASTA version 3.0t78 with the default parameters;

(12) a fusion **protein** comprising a fragment of a first **polypeptide** fused to a second **polypeptide**, the fragment at least 6 consecutive residues of (S1);

(13) producing (V);

(14) identifying (M3) a compound which modulates the activity of a gene product encoded by a nucleic acid comprising (S2), by contacting the gene product with a compound, and determining if the

10/081935

compound modulates the activity of the gene product;

(15) eliciting an immune response in an animal, by introducing into the animal a composition comprising an isolated

polypeptide having 6 consecutive residues of (S1);

(16) a strain of *C. albicans* where a first allele of a gene comprising (S2) is inactive and a second allele is under the control of a heterologous promoter;

(17) identifying a compound or binding partner that binds to a **polypeptide** comprising (S1) or its fragment;

(18) identifying (M4) a compound having the ability to inhibit growth or proliferation of *C. albicans*, by reducing the level or activity of a gene product encoded by (S2) in the *C. albicans* cell relative to a wild type cell, where the reduced level is not lethal to the cell, contacting the cell with a compound and determining if the compound inhibits the growth or proliferation of the cell;

(19) inhibiting (M5) growth or proliferation of *C. albicans* cells, by contacting the cells with the compound that reduce the level of or inhibit the activity of (S1) or (S2);

(20) manufacturing an antimycotic compound, by screening several compounds to identify a compound that reduces the activity or level of a gene product encoded by (S2) and manufacturing the compound identified;

(21) treating (M6) an infection of a subject by *C. albicans*, by administering compound (C) that reduces the activity or level of a gene product encoded by (S2);

(22) an antibody preparation (VI) which binds (V); and

(23) correlating changes in the levels of **proteins** or gene transcripts with the inhibition of growth or proliferation of a diploid fungal cell, by generating a first **protein** expression profile for or transcription profile for a control diploid fungal cell which comprises two wild type alleles of the gene, culturing (I) under conditions where the second allele of the gene is underexpressed, not expressed or overexpressed, and generating a second **protein** expression profile for the cultured cells, and comparing the profiles.

ACTIVITY - Fungicide; Plant Antifungal.

No biological data is given.

MECHANISM OF ACTION - None given.

USE - (M1) is useful for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The diploid fungal cells having both alleles modified are useful for identifying a gene that is essential to the survival or growth of a fungus, a gene that contributes to the virulence and/or pathogenicity of a fungus, a gene that contributes to the resistance of a diploid fungus to an antifungal agent, an antifungal agent that inhibits the growth of a diploid fungus and for identifying a therapeutic agent for treatment of a mammalian disease. (M3) is useful for identifying a compound which modulates the activity of a gene product, preferably enzymatic activity, carbon compound catabolism, biosynthetic, transporter, transcriptional, translational, signal transduction, DNA replication and cell division activity. (M4) is useful for identifying a compound having the ability to inhibit growth or proliferation of *C. albicans*. (M5) is useful for inhibiting growth or proliferation of *C. albicans* cells. (M6) is useful for treating infection by *C. albicans*. (All claimed).

Dwg.0/6

10/081935

L34 ANSWER 5 OF 18 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2002-666824 [71] WPIDS
DOC. NO. CPI: C2002-187137
TITLE: Nucleic acid molecules comprising fungal, e.g.
Cochliobolus heterostrophus, genes from a
peptide synthetase gene cluster, useful for
identifying anti-fungal agents for treating fungal
infections such as pneumonia and arthritis.
DERWENT CLASS: C07 D16
INVENTOR(S): LU, S; TURGEON, B G; YODER, O
PATENT ASSIGNEE(S): (CORR) CORNELL RES FOUND INC; (LUSS-I) LU S; (SYGN)
SYNGENTA PARTICIPATIONS AG; (TURG-I) TURGEON B G;
(YODE-I) YODER O
COUNTRY COUNT: 100
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
WO 2002042444	A2	20020530	(200271)*	EN	316
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZM ZW					
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CO CR CU CZ DE DK DM DZ EC EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ OM PH PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZM ZW					
AU 2002036448	A	20020603	(200277)		

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2002042444	A2	WO 2001-US43381	20011121
AU 2002036448	A	AU 2002-36448	20011121

FILING DETAILS:

PATENT NO	KIND	PATENT NO
AU 2002036448	A Based on	WO 2002042444

PRIORITY APPLN. INFO: US 2000-252732P 20001122; US 2000-252649P
20001122

AN 2002-666824 [71] WPIDS
AB WO 200242444 A UPAB: 20030410
NOVELTY - Nucleic acid molecules comprising fungal, e.g.
Cochliobolus heterostrophus, genes from a **peptide**
synthetase gene cluster, e.g. an iron reductase and/or a permease or
a major facilitator superfamily (MFS) transporter, are new.
DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for
the following:
(1) an isolated polynucleotide (N1) comprising a fungal nucleic
acid segment which encodes a **polypeptide** which is
substantially similar to a **polypeptide**, encoded by a
nucleic acid sequence (NA1) comprising an open reading frame
comprising the 2435 (S46), 1836 (S48), or 2073 (S55) base pair (bp)
sequence defined in the specification, or their complements;
(2) an isolated polynucleotide (N2) comprising a fungal nucleic

acid segment which is substantially similar to NA1, or its complements;

(3) an isolated polynucleotide (N3) comprising a fungal nucleic acid segment which hybridizes under stringent hybridization conditions to S46, S48, S55, or its complements;

(4) an isolated **polypeptide** (P1) encoded by N1, N2 or N3;

(5) an expression cassette comprising a promoter operably linked to N1, N2 or N3;

(6) a **recombinant vector** comprising N1, N2 or N3, where the **vector** is capable of being stably transformed into a host cell;

(7) a host cell comprising the expression cassette of (5) or the **vector** of (6);

(8) a method (M1) for identifying an agent having fungicidal or mycocidal activity, comprising contacting a fungus with an agent that binds to or inhibits P1, and identifying an agent having fungicidal or mycocidal activity;

(9) a method (M2) for identifying an inhibitor of a **polypeptide**, comprising contacting a host cell which expresses a **polypeptide** encoded by N1, N2 or N3 with an agent, and identifying an agent that inhibits the activity of the **polypeptide**;

(10) an agent identified by the M1 or M2;

(11) a method of inhibiting the growth or pathogenicity of a fungus, comprising contacting the fungus with the agent of (10) in an amount sufficient to inhibit the growth or pathogenicity of the fungus;

(12) a method for identifying an agent that modulates a **polypeptide** associated with pathogenicity of a fungus, comprising contacting a fungus with an agent that binds P1, and identifying an agent that modulates the pathogenicity of the fungus;

(13) a method for identifying an agent that modulates the pathogenicity of a fungus, comprising contacting a fungus with an agent that inhibits the activity of P1, and identifying an agent that modulates the pathogenicity of the fungus;

(14) a method (M3) of identifying agents that alter the phenotype of a fungal pathogen or mycogen, comprising:

(a) contacting an agent to be tested with one or more cells of a fungal pathogen or mycogen where:

(i) the cells comprise a nucleotide sequence encoding a **polypeptide** that is substantially similar to the 812 (S47), 611 (S49) or 487 (S56) amino acid sequence defined in the specification; or

(ii) the cells have a mutation in N1, N2 or N3, where the mutation results in overexpression or underexpression of the encoded **polypeptide**; and

(b) detecting or determining whether the agent selectively modulates expression or function or metabolic pathways associated with the **polypeptide**, therefore altering a phenotype of the cells relative to cells not contacted with the agent;

(15) an isolated antibody (Ab1) which specifically binds to P1;

(16) a method (M4) to prevent or inhibit infection of an animal or plant by a fungal pathogen, comprising administering to the animal or plant an effective amount of the agent of (10) for a time and under conditions sufficient to inhibit or prevent fungal growth or reproduction;

(17) the sequence of N1, N2 or N3, or the amino acid sequence

10/081935

of P1 stored on a computer readable medium; and

(18) a transformed plant, the genome of which expresses a chimeric DNA molecule which encodes a gene product which confers resistance or tolerance to the plant to a fungal pathogen by inhibiting fungal iron metabolism or siderophore production.

ACTIVITY - Fungicidal; Mycocidal; Antibacterial; Antiarthritic; Dermatological; Antiinflammatory.

No biological data given.

MECHANISM OF ACTION - Antisense therapy.

No biological data given.

USE - The polynucleotides and **polypeptides** are useful for identifying a novel fungicidal or mycocidal mode of action which permits rapid discovery of novel inhibitors of gene products that are useful as fungicides or mycocides.

The agents and antisense DNA are useful as fungicides to suppress the growth of pathogenic fungi. The fungal pathogens include plant pathogens such as *Septoria tritici*, or *Cochliobolus heterostrophus*, or animal pathogens such as *Candida albicans*.

The anti-fungal agents are particularly useful for treating fungal infections of vertebrates, including immunocompromised vertebrates, for e.g. pneumonia, arthritis, miliary disease, bone and joint infection, skin disease, aesophagitis, vaginitis, onychomycosis, and inflammation of urinary tract, kidney, liver, brain, gastrointestinal tract and lung.

Dwg.0/10

L34 ANSWER 6 OF 18 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN

ACCESSION NUMBER: 2002-739924 [80] WPIDS

DOC. NO. NON-CPI: N2002-582921

DOC. NO. CPI: C2002-209497

TITLE: New **recombinant** spherical outer wall glycoproteins from **Coccidioides immitis**, useful as antigens for serodiagnosis or induction of an immune response.

DERWENT CLASS: B04 C06 C07 D16 S03

INVENTOR(S): COLE, G T; HUNG, C

PATENT ASSIGNEE(S): (MEDI-N) MEDICAL COLLEGE OHIO

COUNTRY COUNT: 1

PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG
US 2002106380	A1	20020808	(200280)*		26

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
US 2002106380	A1	Provisional	
		US 2000-202754P	20000508
		US 2001-850677	20010507

PRIORITY APPLN. INFO: US 2000-202754P 20000508; US 2001-850677 20010507

AN 2002-739924 [80] WPIDS

AB US2002106380 A UPAB: 20021212

NOVELTY - A **recombinant** spherical outer wall glycoprotein 58 (SOWgp58) (I) from **Coccidioides immitis**, is new.

Searcher : Shears 308-4994

10/081935

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for:

- (1) antigens (II) for serodiagnosis of **coccidioidomycosis** comprising a **recombinant** spherical outer wall glycoprotein (SOWgp);
- (2) an isolated nucleic acid (III) that hybridizes under highly stringent conditions to a probe of sequences (S1: 900 base pairs), (S3: 1244 base pairs), (S9: 1504 base pairs) or (S11: 1563 base pairs), or their complements;
- (3) purified **polypeptides** (IV) that comprise sequences (S2: 299 amino acids (aa)), (S4: 328 aa), (S5: 15 aa), (S6: 10 aa), (S7: 10 aa), (S8: 14 aa), (S10: 375 aa) or (S12: 422 aa), all (sic) with up to 20 conservative aa substitutions;
- (4) eliciting (M1) a proliferative response from monocytes, in animals that show a positive skin test, by administering Ag; and
- (5) identifying (M2) the SOW gp58, gp66 and/or gp82 genes.

ACTIVITY - Fungicide.

No biological data given.

MECHANISM OF ACTION - Induction of specific humoral and cellular immune responses.

Purified SOWgp58 from the Silveria isolate was tested in an essentially conventional proliferation assay (thymidine incorporation) against peripheral blood mononuclear cells. At a concentration of 50 micro g/ml in the cell culture medium it produced a stimulation index of 16, using cells from subjects who gave a positive skin test for *C. immitis*, but the index was only about 1 when using cells from subjects who gave a negative skin test.

USE - Spherical outer wall glycoproteins (SOW) are antigens useful for:

- (a) serodiagnosis of **coccidioidomycosis** (San Joaquin Valley fever); and
 - (b) for stimulating proliferation of peripheral blood monocytes, for protection against this disease.
- (III) that encode SOW are useful as probes and primers for cloning, and as diagnostic probes.

Dwg.0/9

L34 ANSWER 7 OF 18 MEDLINE on STN DUPLICATE 2
ACCESSION NUMBER: 2002322692 MEDLINE
DOCUMENT NUMBER: 22060669 PubMed ID: 12065493
TITLE: Role of signal sequence in vaccine-induced protection against experimental **coccidioidomycosis**.
AUTHOR: Jiang Chengyong; Magee D Mitchell; Ivey F Douglas; Cox Rebecca A
CORPORATE SOURCE: Department of Microbiology, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229, USA.
SOURCE: INFECTION AND IMMUNITY, (2002 Jul) 70 (7) 3539-45. Journal code: 0246127. ISSN: 0019-9567.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200207
ENTRY DATE: Entered STN: 20020615
Last Updated on STN: 20020731
Entered Medline: 20020730

Searcher : Shears 308-4994

10/081935

AB The vaccine efficacy of the gene sequence encoding the signal **peptide** of the antigen known as antigen 2 or proline-rich antigen (Ag2/PRA), an immunodominant antigen present in the cell wall of the fungal pathogen **Coccidioides immitis**, was investigated in a murine model of **coccidioidomycosis**. Expression plasmids for Ag2/PRA(1-18) DNA (signal sequence), Ag2/PRA(19-194) DNA (lacking the signal sequence), and Ag2/PRA(1-194) DNA (full length) were inserted in the pVR1012 **vector**, and the constructs were used to vaccinate the highly susceptible BALB/c mouse strain. Immunization with the signal gene sequence significantly reduced the fungal burden in the lungs and spleens of mice 12 days after intraperitoneal challenge with a lethal dose of 2,500 C. immitis arthroconidia, to a level comparable to the protection induced in mice immunized with the full-length Ag2/PRA(1-194) DNA. The Ag2/PRA(19-194) gene protected mice but to a significantly lower level than the signal sequence or the full-length Ag2 gene. The immunizing capacity of Ag2/PRA(1-18) was not attributable to a nonspecific immunostimulatory effect of DNA, as evidenced by the fact that mice immunized with a frameshift mutation of Ag2/PRA(1-18) were not protected against challenge. Furthermore, a synthetic **peptide** corresponding to the translated sequence of Ag2/PRA(1-18) DNA protected mice, albeit at a lower level than the Ag2/PRA(1-18) DNA vaccine. The protection induced with the signal gene vaccine correlated with the production of gamma interferon when splenocytes from Ag2/PRA(1-18)-immunized mice were stimulated with **recombinant** full-length Ag2 and was not associated with the production of anti-**Coccidioides** immunoglobulin G antibody. This is the first study to establish that a signal **peptide** sequence alone, administered as a gene vaccine or synthetic **peptide**, can induce protective immunity against a microbial pathogen.

L34 ANSWER 8 OF 18 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN
ACCESSION NUMBER: 2001-489080 [53] WPIDS
DOC. NO. NON-CPI: N2001-361848
DOC. NO. CPI: C2001-146921
TITLE: Identifying genes essential to fungal metabolisms
and identifying potential therapeutic agents that
target these genes.
DERWENT CLASS: B04 D16 S03
INVENTOR(S): BOONE, C; BUSSEY, H; JIANG, B; ROEMER, T
PATENT ASSIGNEE(S): (ELIT-N) ELITRA PHARM INC
COUNTRY COUNT: 95
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG

WO 2001060975	A2	20010823	(200153)*	EN	320
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC					
MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW					
W: AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA CH CN CR CU CZ DE					
DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG					
KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX MZ NO NZ					
PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN					
YU ZA ZW					
AU 2001043204	A	20010827	(200176)		
EP 1292668	A2	20030319	(200322)	EN	
R: AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK					

10/081935

NL PT RO SE SI TR
KR 2002097180 A 20021231 (200330)
JP 2003523197 W 20030805 (200353) 370

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 2001060975	A2	WO 2001-US5551	20010220
AU 2001043204	A	AU 2001-43204	20010220
EP 1292668	A2	EP 2001-916144	20010220
		WO 2001-US5551	20010220
KR 2002097180	A	KR 2002-710815	20020819
JP 2003523197	W	JP 2001-560347	20010220
		WO 2001-US5551	20010220

FILING DETAILS:

PATENT NO	KIND	PATENT NO
AU 2001043204	A Based on	WO 2001060975
EP 1292668	A2 Based on	WO 2001060975
JP 2003523197	W Based on	WO 2001060975

PRIORITY APPLN. INFO: US 2000-183534P 20000218

AN 2001-489080 [53] WPIDS

AB WO 200160975 A UPAB: 20010919

NOVELTY - Methods for constructing fungal strains useful for identification and validation of gene products as target for therapeutic agents, for identifying and validating gene products as target for therapeutic agents, collections of identified genes and screening assays for the discovery of new drugs, are new.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for the following:

(1) a method (I) of constructing a strain of diploid fungal cells (DFCs) in which both alleles (Als) of a gene are modified and for assembling a collection of DFCs each of which comprises modified Als of a different gene;

(2) a strain of DFCs comprising modified Als of a gene, wherein the first Al of the gene is inactivated by **recombination** using a gene disruption cassette comprising a nucleotide (Nt) sequence (seq) encoding an expressible selectable marker and the expression of the second Al of the gene is regulated by a heterologous promoter that is operably linked to the coding region of the second Al of the gene;

(3) a collection (III) of diploid fungal strains (II) in which each strain comprises modified Als of a different gene, and all the different genes in the genome of the fungus are modified and represented in the collection;

(4) a NAM microarray (IV) comprising a plurality of nucleic acid molecules (NAMs) (each NAM comprises a Nt seq that is hybridizable to a target Nt seq selected from 62 defined seqs given in the specification, a Nt seq that is hybridizable to the Nt seq of a gene that is either essential to the growth of a DFC or contributes to the virulence and/or pathogenicity of the DFCs against a host organism;

(5) a method (V) for identifying a gene that is essential to the metabolism/life cycle of a fungus comprising:

10/081935

(a) culturing the DFCs (II) under conditions in which the second Al of the gene is substantially under expressed (expd) or not expd; and

(b) determining:

(i) viability of the cells (a loss or reduction of viability as compared to a control indicates that the modified gene is essential to the survival of the fungus);

(ii) growth of the cells (a loss or reduction of growth of the cells as compared to a control indicates that the modified gene is essential to the growth of the fungus);

(iii) growth of the cells (a loss or reduction of growth of the cells as compared to a control indicates that the modified gene is essential to the growth of the fungus);

(iv) viability of the cells (an increase in viability as compared to a control indicates that the modified gene contributes to the resistance of the diploid fungus to the antifungal agent);

(6) a method (VI) for identifying an antifungal agent that inhibits the growth of a diploid fungus;

(7) a method (VII) for identifying a therapeutic agent for treatment of a mammalian disease;

(8) a method (VIII) for correlating changes in the levels of **proteins** with the inhibition of growth or proliferation of a DFC, comprising:

(a) generating a first **protein** expression profile for a control DFC which comprises two wild type Als of the gene;

(b) culturing DFCs (II) under conditions wherein the second Al of the gene is substantially under expd, not expd or over expd, and generating a second **protein** expression profile for the cultured cells; and

(c) comparing the first **protein** expression profile with the second **protein** expression profile to identify changes in the levels of **proteins**;

(9) a method (IX) for correlating changes in the levels of gene transcripts with the inhibition of growth or proliferation of a DFC, comprising:

(a) generating a transcription profile for a control DFC which comprises two wild type Als of the gene;

(b) culturing DFCs (II) under conditions wherein the second Al of the gene is substantially under expd, not expd or over expd and generating a second transcription profile for the cultured cells; and

(c) comparing the first transcription profile with the second transcription profile to identify changes in the levels of gene transcripts;

(10) a NAM (X) comprising a Nt seq encoding a gene product required for proliferation of *Candida albicans* (the gene product comprises a defined amino acid seq given in the specification);

(11) a NAM (XI) comprising a fragment of comprising at least 10 to 100 consecutive Nts; a NAM comprising a Nt seq that hybridizes under stringent condition to a second NAM consisting of:

(a) a Nt seq selected from the seqs given in the specification; or

(b) a Nt seq that encodes a **polypeptide** comprising an amino acid seq defined in the specification (the stringent conditions comprises hybridization to filter-bound DNA in 6 multiply sodium chloride sodium citrate (SSC) at about 45 deg. C followed by one or more washes in 0.2 multiply SSC/0.1 % sodium dodecyl sulfate (SDS) at about 50-65 deg. C);

10/081935

(12) a NAM (XII) obtained from an organism other than *Candida albicans* or *Saccharomyces cerevisiae* comprising a Nt seq having at least 30% identity to a defined seq given in the specification, fragments of at least 25 consecutive Nts, and/or seqs complementary to them as determined using BLAST-N version 2.0 with the default parameters;

(13) a **vector** (XIII) comprising a promoter operably linked to the NAMS above;

(14) a host cell (XIV) containing the vector (XIII);

(15) a polypeptide (XV) comprising a defined amino acid seq given in the specification;

(16) a polypeptide (XVI) obtained from an organism other than *Candida albicans* or *Saccharomyces cerevisiae* comprising an amino acid seq having at least 30% similarity to an amino acid seq defined in the specification as determined using FASTA version 3.0t78 with the default parameters;

(17) a fusion protein (XVII) comprising a fragment of a polypeptide fused to a second polypeptide (the fragment comprises at least 6 consecutive residues of a defined amino acid seq given in the specification;

(18) a method (IXX) of producing a polypeptide, comprising introducing into a cell, a vector comprising a promoter operably linked to a Nt seq encoding a polypeptide comprising a defined amino acid seq given in the specification and culturing the cell so that the Nt seq is expd; and

(19) a method (XX) of producing a polypeptide, comprising providing a cell which comprises a heterologous promoter operably linked to a Nt seq encoding a polypeptide comprising a defined amino acid seq given in the specification and culturing the cell so that the Nt seq is expd.

Note: Other INDEPENDENT CLAIMS are included but have had to be omitted due to lack of space.

USE - The methods are used to identify agents that may be used in the treatment of fungal infections.

Dwg.0/6

L34 ANSWER 9 OF 18 MEDLINE on STN DUPLICATE 3
ACCESSION NUMBER: 2001248085 MEDLINE
DOCUMENT NUMBER: 21189198 PubMed ID: 11292702
TITLE: **Recombinant** urease and urease DNA of **Coccidioides immitis** elicit an immunoprotective response against **coccidioidomycosis** in mice.
AUTHOR: Li K; Yu J J; Hung C Y; Lehmann P F; Cole G T
CORPORATE SOURCE: Department of Microbiology and Immunology, Medical College of Ohio, Toledo, Ohio 43614-5806, USA.
CONTRACT NUMBER: AI19149 (NIAID)
AI37232 (NIAID)
SOURCE: INFECTION AND IMMUNITY, (2001 May) 69 (5) 2878-87.
Journal code: 0246127. ISSN: 0019-9567.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 200105
ENTRY DATE: Entered STN: 20010517
Last Updated on STN: 20010517
Entered Medline: 20010510

10/081935

AB **Coccidioides immitis** antigens which stimulate a T helper cell 1 (Th1) pathway of host immune response are considered to be essential components of a vaccine against **coccidioidomycosis**. **Recombinant** urease (rURE) and **recombinant** heat shock **protein** 60 (rHSP60) of *C. immitis* were expressed in *Escherichia coli* and tested as vaccine candidates in BALB/c mice. A synthetic oligodeoxynucleotide which contained unmethylated CpG dinucleotides and was previously shown to enhance a murine Th1 response was used as an immunoadjuvant. T cells isolated from the spleens and lymph nodes of the rURE- and rHSP60-immune mice showed in vitro proliferative responses to the respective **recombinant protein**, but only those T lymphocytes from rURE-immunized mice revealed markedly elevated levels of expression of selected Th1-type cytokine genes. BALB/c mice immunized subcutaneously with rURE and subsequently challenged by the intraperitoneal (i.p.) route with a lethal inoculum of *C. immitis* arthroconidia demonstrated a significant reduction in the level of *C. immitis* infection compared to control animals. rHSP60 was much less effective as a protective antigen. Evaluation of cytokine gene expression in lung tissue and levels of **recombinant** urease-specific immunoglobulins (immunoglobulin G1 [IgG1] versus IgG2a) in murine sera at 12 days after challenge provided additional evidence that immunization with rURE stimulated a Th1 response to the pathogen. Urease was further evaluated by expression of the URE gene in a mammalian plasmid **vector** (pSecTag2A.URE) which was used to immunize mice by the intradermal route. In this case, 82% of the **vector** construct-immunized animals survived more than 40 days after i.p. infection, compared to only 10% of the mice immunized with the **vector** alone. In addition, 87% of the pSecTag2A.URE-immunized survivors had sterile lungs and spleens. These data support the need for further evaluation of the *C. immitis* urease as a candidate vaccine against **coccidioidomycosis**.

L34 ANSWER 10 OF 18 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN

ACCESSION NUMBER: 2002:223315 BIOSIS
DOCUMENT NUMBER: PREV200200223315
TITLE: Replacement of the gene which encodes a major immunoreactive cell surface antigen and adhesin (SOWgp) of the human fungal pathogen **Coccidioides immitis**.
AUTHOR(S): Hung, C. [Reprint author]; Reichard, U. [Reprint author]; Yu, J. [Reprint author]; Cole, G. T. [Reprint author]
CORPORATE SOURCE: Medical College of Ohio, Toledo, OH, USA
SOURCE: Abstracts of the General Meeting of the American Society for Microbiology, (2001) Vol. 101, pp. 365. print.
Meeting Info.: 101st General Meeting of the American Society for Microbiology. Orlando, FL, USA. May 20-24, 2001. American Society of Microbiology. ISSN: 1060-2011.
DOCUMENT TYPE: Conference; (Meeting)
Conference; Abstract; (Meeting Abstract)
LANGUAGE: English
ENTRY DATE: Entered STN: 3 Apr 2002
Last Updated on STN: 3 Apr 2002

Searcher : Shears 308-4994

10/081935

AB We have previously cloned and expressed a major parasitic cell surface antigen (SOWgp) of *C. immitis* which, upon immunization elicits both antibody and T cell responses in BALB/c mice. Our data also showed that the **recombinant** SOWgp has high affinity for laminin and fibronectin of the extracellular matrix (ECM). To confirm that the SOWgp functions as an adhesin, the SOWgp gene was deleted by a double, locus-specific crossover event. This resulted in replacement of 1.8-kb fragment of the SOWgp gene with a 3.6-kb DNA construct that encodes a hygromycin resistant gene from the pAN7.1 **vector**. Results of PCR analysis, Southern hybridization, and immunoblot assays of parasitic cell homogenates of the Dsowgp mutant compared to the parental strain confirmed that the SOWgp gene was replaced and the antigen was no longer produced. This is the first report of a successful targeted gene knockout of *C. immitis* using the double crossover strategy. Comparative immunoelectron-microscopic examinations of the isolated, crude outer parasitic cell wall fraction of the DELTAsowgp and parental strains showed total loss of patient antibody reactivity with the former. The DELTAsowgp parasitic cells showed 20-50% reduction in their ability to bind to human and murine ECM. Virulence studies of the DELTAsowgp strain in intranasally-challenged BALB/c mice showed 20% survival compared to none in the case of mice challenged with the parental strain. It is apparent that deletion of SOWgp production did not eliminate the ability of the pathogen to cause disease. Preliminary investigations of the influence of **recombinant** SOWgp on immune murine T cells in vitro suggested that the **protein** stimulates a dominant Th2 response, demonstrated by elevated levels of IL-5 production. Our hypothesis is that SOWgp has a major impact on disseminated **coccidioidal** infection by compromising the host immune pathway toward a detrimental Th2 response.

L34 ANSWER 11 OF 18 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN

ACCESSION NUMBER: 2002:223274 BIOSIS
DOCUMENT NUMBER: PREV200200223274
TITLE: Isolation and characterization of a macrophage binding **protein** of **Coccidioides immitis**.
AUTHOR(S): Cole, G. T. [Reprint author]; Fradin, C. [Reprint author]
CORPORATE SOURCE: Medical College of Ohio, Toledo, OH, USA
SOURCE: Abstracts of the General Meeting of the American Society for Microbiology, (2001) Vol. 101, pp. 356. print.
Meeting Info.: 101st General Meeting of the American Society for Microbiology. Orlando, FL, USA. May 20-24, 2001. American Society of Microbiology. ISSN: 1060-2011.
DOCUMENT TYPE: Conference; (Meeting)
Conference; Abstract; (Meeting Abstract)
LANGUAGE: English
ENTRY DATE: Entered STN: 3 Apr 2002
Last Updated on STN: 3 Apr 2002

AB Early studies have demonstrated the significance of the interplay between **Coccidioides immitis** and macrophages during **coccidioidomycosis**. In this study we report the purification of a ligand expressed at the surface of endospores of

10/081935

C. immitis which binds to an unknown receptor of a murine macrophage line (IC-21). Cell wall **proteins** of different stages of the parasitic cycle were labeled with biotin and subjected to beta-mercaptoethanol extraction. A single macrophage plasma membrane binding **protein** (MBP) with molecular size of 36-kDa was isolated from the endospore extract. The C. immitis MBP was purified and two internal **peptides** of the trypsin-digested fraction were sequenced. Degenerate primers were designed to PCR amplify a fragment of the MBP gene, and the PCR product was then used to screen a C. immitis genomic library for isolation of the full-length gene. The MBP gene is characterized by an open reading frame of 638 bp, and does not contain any introns. The MBP gene was subcloned into the pET28b expression **vector** and the **recombinant**, histidine-tagged fusion **protein** (rMBP) was produced by transformed E. coli strain BL21(DE3). The rMBP was shown to bind to the surface of both IC-21 cells and resident murine peritoneal macrophages. An in vitro study of temporal expression of the MBP gene during the parasitic cycle was conducted by RT-PCR and Northern blot analyses, and showed that maximum expression occurs at the endospore stage. Polyclonal antibody raised in guinea pigs against the rMBP was used for immunoblot analyses of beta-mercaptoethanol extracts of intact parasitic cells, as well as total cell homogenates of the saprobic and parasitic phases of C. immitis. The results showed that the rMBP was present only in beta-mercaptoethanol extracts and homogenates of first generation round cells and endospores. We suggest that the newly-isolated MBP of C. immitis plays a pivotal role in the course of **coccidioid** infections:

L34 ANSWER 12 OF 18 WPIDS COPYRIGHT 2003 THOMSON DERWENT on STN .
ACCESSION NUMBER: 1999-551417 [46] WPIDS
CROSS REFERENCE: 2002-606631 [65]; 2003-829776 [77]
DOC. NO. NON-CPI: N1999-407989
DOC. NO. CPI: C1999-161024
TITLE: Novel chitin-binding fragments of human chitinase
used to treat fungal infections in animals.
DERWENT CLASS: B04 C06 D16 S03
INVENTOR(S): GRAY, P W; TJOELKER, L W
PATENT ASSIGNEE(S): (ICOS-N) ICOS CORP
COUNTRY COUNT: 86
PATENT INFORMATION:

PATENT NO	KIND	DATE	WEEK	LA	PG

WO 9946390	A1	19990916	(199946)*	EN	83
RW: AT BE CH CY DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC					
MW NL OA PT SD SE SL SZ UG ZW					
W: AE AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES					
FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK					
LR LS LT LU LV MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG					
SI SK SL TJ TM TR TT UA UG UZ VN YU ZW					
AU 9929989	A	19990927	(200006)		
BR 9908724	A	20001121	(200065)		
NO 2000004522	A	20001101	(200065)		
EP 1078073	A1	20010228	(200113)	EN	
R: AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE					
US 6200951	B1	20010313	(200120)		
CZ 2000003308	A3	20010711	(200147)		

Searcher : Shears 308-4994

10/081935

HU 2001001078 A2 20010730 (200157)
SK 2000001354 A3 20010710 (200157)
MX 2000008863 A1 20010301 (200170)
JP 2002505882 W 20020226 (200219) 86
CN 1357046 A 20020703 (200265)
AU 763582 B 20030724 (200355)

APPLICATION DETAILS:

PATENT NO	KIND	APPLICATION	DATE
WO 9946390	A1	WO 1999-US5343	19990312
AU 9929989	A	AU 1999-29989	19990312
BR 9908724	A	BR 1999-8724	19990312
		WO 1999-US5343	19990312
NO 2000004522	A	WO 1999-US5343	19990312
		NO 2000-4522	20000911
EP 1078073	A1	EP 1999-911320	19990312
		WO 1999-US5343	19990312
US 6200951	B1	US 1998-39198	19980312
CZ 2000003308	A3	WO 1999-US5343	19990312
		CZ 2000-3308	19990312
HU 2001001078	A2	WO 1999-US5343	19990312
		HU 2001-1078	19990312
SK 2000001354	A3	WO 1999-US5343	19990312
		SK 2000-1354	19990312
MX 2000008863	A1	MX 2000-8863	20000911
JP 2002505882	W	WO 1999-US5343	19990312
		JP 2000-535757	19990312
CN 1357046	A	CN 1999-805990	19990312
AU 763582	B	AU 1999-29989	19990312

FILING DETAILS:

PATENT NO	KIND	PATENT NO
AU 9929989	A Based on	WO 9946390
BR 9908724	A Based on	WO 9946390
EP 1078073	A1 Based on	WO 9946390
CZ 2000003308	A3 Based on	WO 9946390
HU 2001001078	A2 Based on	WO 9946390
SK 2000001354	A3 Based on	WO 9946390
JP 2002505882	W Based on	WO 9946390
AU 763582	B Previous Publ.	AU 9929989
	Based on	WO 9946390

PRIORITY APPLN. INFO: US 1998-39198 19980312

AN 1999-551417 [46] WPIDS
CR 2002-606631 [65]; 2003-829776 [77]
AB WO 9946390 A UPAB: 20031128

NOVELTY - Novel human chitinase **polypeptide** and polynucleotide fragments have chitin-binding activity but lack chitinase enzymatic activity.

DETAILED DESCRIPTION - A chitin-binding, chitinase-inactive **polypeptide** (I) comprises a chitin-binding fragment of the 54 C-terminal amino acids (aa) of human chitinase having a 466 aa sequence (S1) (given in the specification).

INDEPENDENT CLAIMS are also included for the following:

10/081935

- (1) a chitin-binding, chitinase-inactive **polypeptide** comprising aa residues X-Y of (S1), where X is a consecutive integer from 347 to 397 and Y is 445;
- (2) a fusion **protein** comprising (I) fused to a heterologous **polypeptide** (e.g. an enzyme);
- (3) a composition comprising (I) and a diluent, optionally further comprising a non-chitinase anti-fungal agent;
- (4) a composition comprising (I) or the fusion **protein** of (1) conjugated to an anti-fungal agent;
- (5) treating fungal infection, comprising administering to a subject suffering from fungal infection the composition of (3) or (4), optionally further comprising administering a non-chitinase anti-fungal agent;
- (6) a composition comprising (I) or the fusion **protein** of (1) conjugated to a detectable label, especially radioisotopes, fluorophores, dyes, electron-dense compounds or enzymes;
- (7) determining the presence of chitin in a sample, comprising contacting the sample with the composition of (6), and determining the amount of labelled **polypeptide** bound to chitin;
- (8) a purified, isolated polynucleotide (especially DNA) encoding (I);
- (9) a **vector** comprising the DNA of (8);
- (10) a host cell stably transformed or transfected with the DNA of (8) in a manner allowing the expression of (I), and
- (11) a monoclonal antibody that specifically binds to an epitope within the 54 C-terminal amino acids of human chitinase having a sequence (S1).

ACTIVITY - Chitin-binding.

MECHANISM OF ACTION - None given.

USE - Chitinase fragments can be used to screen for **proteins** or other molecules that specifically bind to the chitin-binding domain of human chitinase or that modulate its activity. These compounds are useful for immunization, as well as for purifying chitinase, as well as for detection and quantification of chitinase. Polynucleotide fragments of the invention are useful as a source of probes and primers, and to express the **proteins recombinantly**. The chitinase fragments, when conjugated to antifungal compounds, are used to treat animals, especially humans, infected with chitin-containing parasites such as fungi. Fungal infection treated include candidiasis, aspergillosis, **coccidioidomycosis**, blastomycosis, paracoccidioidomycosis, mucormycosis, histoplasmosis, cryptococcosis, chromoblastomycosis, sporotrichosis, and dermatophytoses.

ADVANTAGE - Chitin can be degraded by the enzyme chitinase. Use of whole chitinase **protein** for treating infections, especially fungal infections, is problematic. In view of the increasing incidences of life-threatening fungal infection in e.g. immunocompromised individuals, there exists a need for identifying new compounds for treating fungal infection. The chitin-binding fragments of the present invention provide this need.
Dwg.0/0

L34 ANSWER 13 OF 18

MEDLINE on STN

DUPLICATE 4

ACCESSION NUMBER: 1999085554 MEDLINE

DOCUMENT NUMBER: 99085554 PubMed ID: 9868663

TITLE: Detecting serum antibodies to a purified
recombinant proline-rich antigen of
Coccidioides immitis in patients with

10/081935

coccidioidomycosis.
AUTHOR: Orsborn K I; Galgiani J N
CORPORATE SOURCE: Research and Medical Services, Veterans Affairs
Medical Center, Tucson, Arizona, USA.
SOURCE: CLINICAL INFECTIOUS DISEASES, (1998 Dec) 27 (6)
1475-8.
Journal code: 9203213. ISSN: 1058-4838.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199903
ENTRY DATE: Entered STN: 19990324
Last Updated on STN: 19990324
Entered Medline: 19990308

AB In previous work, antibodies in serum samples from patients with **coccidioidomycosis** were found to react with a proline-rich antigen (PRA) isolated from spherules of **Coccidioides immitis**, and the gene encoding this antigen was cloned. We expressed and purified **recombinant** PRA (rPRA) by removing the majority of amino acids contributed by the **vector** from the fusion **protein**. Purified rPRA reacted with serum IgG antibodies in 37 of 42 patients with culture-proven progressive pulmonary or extrapulmonary **coccidioidal** disease; specific antibodies in dilutions ranging from 1:40 to 1:102,400 were demonstrated (sensitivity, 88%). In contrast, for > 95% of patients without **coccidioidomycosis** reactivity of < 1:40 was demonstrated (specificity, 97%). Of 18 patients with primary self-limited **coccidioidomycosis**, none had detectable antibodies in serum samples collected up to 141 days after illness began. The association of antibodies to rPRA with progressive infection may have prognostic value.

L34 ANSWER 14 OF 18 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN

ACCESSION NUMBER: 1997:500830 BIOSIS
DOCUMENT NUMBER: PREV199799800033
TITLE: Isolation and characterization of the urease gene (URE) from the pathogenic fungus **Coccidioides immitis**.
AUTHOR(S): Yu, Jieh-Juen; Smithson, S. L.; Thomas, Pei W.; Kirkland, Theo N.; Cole, Garry T. [Reprint author]
CORPORATE SOURCE: Dep. Microbiol. Immunol., Med. Coll. Ohio, 3000 Arlington Ave., Toledo, OH 43699, USA
SOURCE: Gene (Amsterdam), (1997) Vol. 198, No. 1-2, pp. 387-391.
CODEN: GENED6. ISSN: 0378-1119.
DOCUMENT TYPE: Article
LANGUAGE: English
ENTRY DATE: Entered STN: 21 Nov 1997
Last Updated on STN: 21 Nov 1997

AB The urease (URE)-encoding gene from **Coccidioides immitis** (Ci), a respiratory fungal pathogen of humans, was cloned, sequenced, chromosome-mapped and expressed. Both the genomic and cDNA sequences are reported. The transcription start point and poly(A)-addition site were confirmed. The URE gene contains eight introns and a 2517-bp ORF that translates a 839-amino-acid (aa) **protein** of 91.5 kDa and pI of 5.5, as deduced by computer

10/081935

analysis of the nucleotide sequence. The translated **protein** revealed eight putative N-glycosylation sites. The deduced URE showed comparable levels of homology to reported URE of the jack bean plant (*Canavalia ensiformis*; 71.8%) and URE of several genera of bacteria (Bp, 71.7%; Hp, 68.3%; Ka, 71.6%; Pm, 71.9%). The URE gene was mapped to chromosome III of Ci and was shown to be a single copy gene by Southern hybridization. Expression of a 1687-bp fragment of the URE gene in *E. coli* resulted in the production of a 63-kDa **recombinant protein** that was recognized in an immunoblot by antiserum raised against the Ka URE homolog. This is the first report of a fungal URE gene.

L34 ANSWER 15 OF 18 MEDLINE on STN DUPLICATE 5
ACCESSION NUMBER: 96294782 MEDLINE
DOCUMENT NUMBER: 96294782 PubMed ID: 8698497
TITLE: Molecular cloning and characterization of
Coccidioides immitis antigen 2 cDNA.
AUTHOR: Zhu Y; Yang C; Magee D M; Cox R A
CORPORATE SOURCE: Department of Clinical Investigation, Texas Center
for Infectious Disease, San Antonio 78223, USA.
CONTRACT NUMBER: AI2143 (NIAID)
SOURCE: INFECTION AND IMMUNITY, (1996 Jul) 64 (7) 2695-9.
Journal code: 0246127. ISSN: 0019-9567.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
OTHER SOURCE: GENBANK-U32518
ENTRY MONTH: 199609
ENTRY DATE: Entered STN: 19960912
Last Updated on STN: 19960912
Entered Medline: 19960904

AB Previous experiments have provided evidence that **Coccidioides immitis** antigen 2 (Ag2) is a major T-cell-reactive component of mycelia and spherule cell walls. Here we report the identification and cloning of the cDNA that encodes Ag2 from a lambda ZAP cDNA expression library constructed from spherule-derived RNA. DNA sequence analysis established that the 1,255-bp clone contains a 174-bp 5' untranslated region, a 582-bp open reading frame which encodes for a **protein** consisting of 194 amino acids, and a 375-bp 3' untranslated region, including a poly(A) tail. The **recombinant Ag2 protein** has a predicted molecular mass of 19.5 kDa and contains an 18-amino-acid N terminus which has been tentatively identified as a signal **peptide**. The Ag2 cDNA was ligated into the pGEX-4T-3 **vector** and expressed in *Escherichia coli* TG-1 cells as a glutathione S-transferase fusion **protein**. The **recombinant fusion protein** showed reactivity with sera from patients with **coccidioidomycosis** and elicited delayed-type footpad hypersensitivity responses in **Coccidioides**-immune mice. These results suggest that the Ag2 cDNA can be used for the large-scale production of this immunologically important **protein**.

L34 ANSWER 16 OF 18 MEDLINE on STN DUPLICATE 6
ACCESSION NUMBER: 96239008 MEDLINE
DOCUMENT NUMBER: 96239008 PubMed ID: 8675298
TITLE: Molecular cloning and characterization of the

10/081935

Coccidioides immitis complement
fixation/chitinase antigen.
AUTHOR: Yang C; Zhu Y; Magee D M; Cox R A
CORPORATE SOURCE: Department of Clinical Investigation, Texas Center
for Infectious Diseases, San Antonio, 78223, USA.
CONTRACT NUMBER: AI 12431 (NIAID)
SOURCE: INFECTION AND IMMUNITY, (1996 Jun) 64 (6) 1992-7.
Journal code: 0246127. ISSN: 0019-9567.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
OTHER SOURCE: GENBANK-U33265
ENTRY MONTH: 199608
ENTRY DATE: Entered STN: 19960822
Last Updated on STN: 19960919
Entered Medline: 19960812

AB Detection of anti-**Coccidioides** complement-fixing (CF)
antibody is a valuable diagnostic and prognostic aid in
coccidioidomycosis. The CF antibody response is directed
against a heat-labile antigen that has chitinase activity, hereafter
referred to as the CF/chitinase **protein**. To identify and
clone this immunoreactive enzyme, we constructed a
Coccidioides immitis cDNA lambda ZAP expression library from
spherule RNA and detected fusion **peptides** expressing CF
epitopes by immunoscreening. A cDNA clone consisting of 1,623 bp
was identified, sequenced, and found to contain a single open
reading frame that encodes a **protein** of 47 kDa with 427
amino acids. Deduced amino acid sequence analyses showed that the
cloned CF/chitinase cDNA contains a 35-amino-acid region, beginning
at Ser-18 and ending at and ending at Arg-52 which has 92% homology
with the reported N-terminal amino acid sequence of authentic
CF/chitinase **protein**. The first 17 amino acids in the
deduced sequence of the cloned cDNA are not present on the mature
CF/chitinase **protein**, suggesting that it may be a signal
peptide. Expression of the CF/chitinase cDNA insert by
using the pGEX-4T-3 **vector** yields a fusion **peptide**
that bears CF-specific epitopes and shows chitinase activity. The
CF/chitinase clone will enable large-scale production of the
recombinant CF antigen for use in immunoassays and
facilitate studies on the role of chitinase in the morphogenesis of
C. immitis.

L34 ANSWER 17 OF 18 MEDLINE on STN DUPLICATE 7
ACCESSION NUMBER: 96009757 MEDLINE
DOCUMENT NUMBER: 96009757 PubMed ID: 7558310
TITLE: Molecular and biochemical characterization of a
Coccidioides immitis-specific antigen.
AUTHOR: Pan S; Cole G T
CORPORATE SOURCE: Department of Microbiology, Medical College of Ohio,
Toledo 43699-0008, USA.
CONTRACT NUMBER: AI19149 (NIAID)
SOURCE: INFECTION AND IMMUNITY, (1995 Oct) 63 (10) 3994-4002.
Journal code: 0246127. ISSN: 0019-9567.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals

Searcher : Shears 308-4994

10/081935

OTHER SOURCE: GENBANK-L36551
ENTRY MONTH: 199510
ENTRY DATE: Entered STN: 19951227
Last Updated on STN: 19951227
Entered Medline: 19951030

AB Results of earlier investigations have indicated that the saprobic phase of **Coccidioides immitis** produces a heat-stable, 19-kDa antigen with serine proteinase activity which has been suggested to be specific for this pathogenic fungus. In the present study we have determined the N-terminal and partial internal amino acid sequences of the purified, 19-kDa antigen, cloned the gene which encodes this **polypeptide**, and confirmed that the secreted proteinase is a **Coccidioides**-specific antigen (CS-Ag). Both the genomic and cDNA sequences are reported and reveal that the *csa* gene which encodes this antigen has no introns. A 543-bp open reading frame encodes a 181-amino-acid-containing **protein** with a predicted molecular mass of 19.8 kDa and an isoelectric point of 8.3. The *csa* gene was localized on chromosome I of three representative *C. immitis* clinical isolates on the basis of Southern hybridizations. Expression of the *csa* gene in *Escherichia coli* using the pET21a plasmid **vector** yielded a **recombinant protein** that was recognized in immunoblot assays by antibody raised to the purified 19-kDa CS-Ag. Secretion of the native antigen is suggested to occur by cleavage of a putative 23-residue signal **peptide**. The native CS-Ag showed a low degree of glycosylation. Analysis of the carbohydrate composition of the CS-Ag revealed xylose, mannose, galactose, and glucose. However, the purified antigen showed no affinity for concanavalin A. A PCR method with specificity and high sensitivity for detection of *C. immitis* genomic DNA, using a pair of synthetic oligonucleotide primers whose sequences were based on that of the *csa* gene, was developed. A 520-bp product was amplified only when *C. immitis* genomic DNA was used as the template. The lower limits of DNA detection using this PCR method were 1 pg of *C. immitis* genomic DNA by ethidium bromide staining and 100 fg after Southern hybridization. The *csa* gene-based PCR method for detection of *C. immitis* DNA is useful for culture identification and may have clinical applications for the diagnosis of **coccidioidal** infections.

L34 ANSWER 18 OF 18 MEDLINE on STN
ACCESSION NUMBER: 94129182 MEDLINE
DOCUMENT NUMBER: 94129182 PubMed ID: 8298279
TITLE: Identification of antigens of **Coccidioides immitis** which stimulated immune T lymphocytes.
AUTHOR: Cole G T; Kirkland T N
CORPORATE SOURCE: Department of Botany, University of Texas, Austin 78713.
CONTRACT NUMBER: AI19149 (NIAID)
SOURCE: ARCHIVES OF MEDICAL RESEARCH, (1993 Autumn) 24 (3) 281-91.
Journal code: 9312706. ISSN: 0188-4409.
PUB. COUNTRY: Mexico
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199403
ENTRY DATE: Entered STN: 19940318

10/081935

Last Updated on STN: 19940318

Entered Medline: 19940308

AB T-cell mediated immune response to **coccidioidomycosis** has been shown to be the principal mechanism of resistance to this respiratory fungal disease in experimental animals. In this study, a **Coccidioides immitis** antigen-specific murine T-cell line was used to identify macromolecules capable of eliciting an immune mouse T-cell proliferative response. The murine T-cell line was selected on the basis of its strong positive response to a soluble conidial wall fraction (SCWF), which had previously been shown to be reactive in humoral and cellular immunoassays. An antigen-specific T-cell line rather than T-cell clones was used to identify multiple antigens. The T-cell immunoblot method was employed first to identify immunoreactive sub-fractions of the SCWF, and then to identify T-cell fusion **proteins** (FPs) obtained from a cDNA expression library constructed in lambda gt11. The library was screened with anti-SCWF. The nucleotide sequence of a 0.2 kilobase cDNA insert encoding a FP which elicited vigorous T-cell response was determined. A construct of this insert was subcloned into the pET expression **vector** system and 6.5-kilodalton (kDa) **recombinant protein** (RP) expressed in *Escherichia coli* was isolated. The RP and FP were shown to be homologous on the basis of identify of their amino acid sequences. Antibody raised in guinea pigs against the RP recognized a 59-kDa native **protein** of the mycelial culture filtrate produced by three separate strains of *C. immitis*, and reacted with the cell wall of arthroconidia as detected by immunofluorescence microscopy. In this study we have identified and partly characterized a potentially important T-cell stimulating antigen of *C. immitis*.

(FILE 'MEDLINE' ENTERED AT 09:35:53 ON 04 DEC 2003)

L35 720 SEA FILE=MEDLINE ABB=ON PLU=ON COCCIDIOIDES/CT
L36 120410 SEA FILE=MEDLINE ABB=ON PLU=ON PROTEINS/CT
L37 86 SEA FILE=MEDLINE ABB=ON PLU=ON POLYPROTEINS/CT
L38 73063 SEA FILE=MEDLINE ABB=ON PLU=ON PEPTIDES/CT
L39 7 SEA FILE=MEDLINE ABB=ON PLU=ON L35 AND (L36 OR L37 OR L38)

L39 ANSWER 1 OF 7 MEDLINE on STN

AN 2002322678 MEDLINE

TI Localization within a proline-rich antigen (Ag2/PRA) of protective antigenicity against infection with *Coccidioides immitis* in mice.

AU Peng Tao; Shubitz Lisa; Simons Julie; Perrill Robert; Orsborn Kris I; Galgiani John N

SO INFECTION AND IMMUNITY, (2002 Jul) 70 (7) 3330-5.

Journal code: 0246127. ISSN: 0019-9567.

AB Subunits of a proline-rich coccidioidal antigen (Ag2/PRA) of *Coccidioides immitis* were analyzed by comparison as vaccines in mice. The optimal dose of plasmid vaccine encoding full-length Ag2/PRA was determined to be between 10 and 100 microg. Mice vaccinated with plasmids encoding amino acids (aa) 1 to 106 were as protective as full-length Ag2/PRA (aa 1 to 194). The subunit from aa 27 to 106 was significantly but less protective. Plasmids encoding aa 90 to 151 or aa 90 to 194 were not protective. Analogous results were obtained with recombinant vaccines of the same amino acid sequences. In addition, mixtures of aa 90 to 194 with either aa 1 to 106 or aa 27 to 106 did not enhance protection compared to the active single-recombinant subunits alone. Humoral

10/081935

response of total immunoglobulin G (IgG) and subclasses IgG1 and IgG2a were detectable in subunit vaccinations but at significantly (100-fold) lower concentrations than after vaccination with plasmids encoding full-length Ag2/PRA. Since virtually all protection by vaccination with full-length Ag2/PRA can be accounted for in the first half of the protein (aa 1 to 106), this subunit could make a multicomponent vaccine more feasible by reducing the quantity of protein per dose and the possibility of an untoward reactions to a foreign protein.

- L39 ANSWER 2 OF 7 MEDLINE on STN
AN 1999094977 MEDLINE
TI Proline-rich vaccine candidate antigen of *Coccidioides immitis*: conservation among isolates and differential expression with spherule maturation.
AU Peng T; Orsborn K I; Orbach M J; Galgiani J N
SO JOURNAL OF INFECTIOUS DISEASES, (1999 Feb) 179 (2) 518-21.
Journal code: 0413675. ISSN: 0022-1899.
AB A proline rich antigen (PRA), which protects mice against *Coccidioides immitis*, has been analyzed for differential expression and variation among isolates. Northern blots of RNA from different stages of growth were probed with previously cloned cDNA and showed that mRNA for PRA increased as spherules transformed and matured from mycelia. This pattern corresponds to the relative potency of whole cell vaccines from similar preparations. The PRA gene was then cloned from a genomic library of the Silveira strain of *C. immitis* and its sequence analyzed. Eight other coccidioidal isolates, selected for diversity in geographic origin and resulting clinical disease, demonstrated heterogeneity in Southern blots and in sequences of polymerase chain reaction products. Silveira differed from other California isolates at 33 of 555 bases, whereas it differed from non-California isolates by ≤ 2 bases. Only one of these substitutions affected protein sequence. Thus, tests or vaccines based on this gene are likely to cover most isolates.
- L39 ANSWER 3 OF 7 MEDLINE on STN
AN 96144702 MEDLINE
TI Cloning and sequence analysis of the cDNA for a protein from *Coccidioides immitis* with immunogenic potential.
AU Dugger K O; Villareal K M; Ngyuen A; Zimmermann C R; Law J H; Galgiani J N
SO BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, (1996 Jan 17) 218 (2) 485-9.
Journal code: 0372516. ISSN: 0006-291X.
AB We have cloned and sequenced the cDNA encoding an immunoreactive protein from the pathogenic fungus *Coccidioides immitis* which stimulates human T cells and has been associated with protective vaccines in mice. The transcript contained an open reading frame encoding 194 amino acids with a calculated molecular weight of 19.5 kDa, a 151 base 5' untranslated region (UTR), and a 468 base 3'UTR. A four member repeat motif, usually thr-ala-glu-pro, exists for amino acids 98 through 141. Deduced amino acid sequence derived from the cDNA was identical with previously determined internal amino acid sequence from the native protein, and goat antiserum raised against the purified fungal protein reacted with an inducible fusion protein translated from this cDNA. Using this cDNA to produce recombinant protein will allow direct testing of its role in human immunity to coccidioidomycosis and may lead to new diagnostic

10/081935

tests.

L39 ANSWER 4 OF 7 MEDLINE on STN
AN 90383461 MEDLINE
TI Activity of cilofungin against *Coccidioides immitis*: differential in vitro effects on mycelia and spherules correlated with in vivo studies.
AU Galgiani J N; Sun S H; Clemons K V; Stevens D A
SO JOURNAL OF INFECTIOUS DISEASES, (1990 Oct) 162 (4) 944-8.
Journal code: 0413675. ISSN: 0022-1899.
AB Cilofungin, a new antifungal drug, was found to inhibit mycelial growth of *Coccidioides immitis*. Light and electron microscopic observations indicated delay in development of the outer hyphal wall. Cilofungin also blocked incorporation of the chitin substrate, N-acetylglucosamine, into mycelia. However, when *C. immitis* was grown under conditions that induced spherule development, drug effects were dramatically decreased. Furthermore, efficacy of cilofungin in treatment of murine coccidioidomycosis could not be demonstrated. These studies indicate that glucan-synthase inhibitors have activity against *C. immitis*, and other compounds with different pharmacologic properties or in combination with other antifungal drugs may exploit this biologic effect.

L39 ANSWER 5 OF 7 MEDLINE on STN
AN 75150364 MEDLINE
TI Immunocompetence and prognosis in patients with gynecologic cancer.
AU Nalick R H; Disaia P J; Rea T H; Morrow C P
SO GYNECOLOGIC ONCOLOGY, (1974 Apr) 2 (1) 81-92.
Journal code: 0365304. ISSN: 0090-8258.

L39 ANSWER 6 OF 7 MEDLINE on STN
AN 72041293 MEDLINE
TI Fractionation and composition studies of skin test-active components of sensitins from *Coccidioides immitis*.
AU Anderson K L; Wheat R W; Conant N F
SO APPLIED MICROBIOLOGY, (1971 Sep) 22 (3) 294-9.
Journal code: 7605802. ISSN: 0003-6919.

L39 ANSWER 7 OF 7 MEDLINE on STN
AN 68239308 MEDLINE
TI Sarcoidosis. II. A clinical evaluation of the alteration in delayed hypersensitivity.
AU Lordon R E; Young R L; Shapiro S S; Smith R E; Weg J G
SO AMERICAN REVIEW OF RESPIRATORY DISEASE, (1968 Jun) 97 (6) 1009-16.
Journal code: 0370523. ISSN: 0003-0805.

(FILE 'HCAPLUS' ENTERED AT 09:41:53 ON 04 DEC 2003)

L52 208 SEA FILE=HCAPLUS ABB=ON PLU=ON C IMMITIS
L53 74 SEA FILE=HCAPLUS ABB=ON PLU=ON L52 AND (PROTEIN OR POLYPEPTIDE OR PROTEIN OR PEPTIDE)
L54 10 SEA FILE=HCAPLUS ABB=ON PLU=ON L53 AND VECTOR
L55 7 SEA FILE=HCAPLUS ABB=ON PLU=ON L54 AND RECOMBIN?

L56 0 L55 NOT L32

(FILE 'MEDLINE, BIOSIS, EMBASE, WPIDS, CONFSCI, SCISEARCH,

10/081935

JICST-EPLUS, JAPIO' ENTERED AT 09:43:13 ON 04 DEC 2003)

L57 22 S L55
L58 0 S L57 NOT L33

(FILE 'USPATFULL' ENTERED AT 09:44:23 ON 04 DEC 2003)

L59 635 SEA FILE=USPATFULL ABB=ON PLU=ON (COCCIDIOD? OR
C) (W) IMMITIS OR COCCIDIOIDOMYC?
L60 4 SEA FILE=USPATFULL ABB=ON PLU=ON L59(10A) (PROTEIN OR
PEPTIDE OR POLYPEPTIDE OR POLYPROTEIN)

L59 635 SEA FILE=USPATFULL ABB=ON PLU=ON (COCCIDIOD? OR
C) (W) IMMITIS OR COCCIDIOIDOMYC?
L65 246 SEA FILE=USPATFULL ABB=ON PLU=ON L59(S) (PROTEIN OR
PEPTIDE OR POLYPEPTIDE OR POLYPROTEIN)
L66 2 SEA FILE=USPATFULL ABB=ON PLU=ON L65(S) (VECTOR(3A) RECOM
BIN?)

L67 6 L60 OR L66

L67 ANSWER 1 OF 6 USPATFULL on STN

ACCESSION NUMBER: 2003:311860 USPATFULL
TITLE: Glucanoyl transferase-1 protein useful for
immunization against Coccidioides spp.
INVENTOR(S): Cole, Garry T., Toledo, OH, UNITED STATES
Delgado, Nelson, Delran, NJ, UNITED STATES
Yu, Jieh-Juen, Toledo, OH, UNITED STATES
Xue, Jianmin, Toledo, OH, UNITED STATES

	NUMBER	KIND	DATE
PATENT INFORMATION:	US 2003219455	A1	20031127
APPLICATION INFO.:	US 2003-417997	A1	20030416 (10)

	NUMBER	DATE
PRIORITY INFORMATION:	US 2002-374152P	20020422 (60)
DOCUMENT TYPE:	Utility	
FILE SEGMENT:	APPLICATION	
LEGAL REPRESENTATIVE:	Michael R. Ward, Morrison & Foerster LLP, 425 Market Street, San Francisco, CA, 94105-2482	
NUMBER OF CLAIMS:	52	
EXEMPLARY CLAIM:	1	
NUMBER OF DRAWINGS:	4 Drawing Page(s)	
LINE COUNT:	2165	

AB The present invention provides compositions of
1,3- β -glucanoyltransferase (Gel-1) antigens and
polynucleotides encoding the Gel-1 antigens, which Gel-1 antigens
are useful for generating an immunological response in an
individual and in therapeutic and diagnostic applications of
infections due to pathogenic Coccidioides spp. fungi, such as C.
immitis or C. posadasii.

INCL INCLM: 424/190.100
INCLS: 435/069.300; 435/070.210; 435/252.300; 435/320.100;
530/350.000; 536/023.700

Searcher : Shears 308-4994

10/081935

NCL NCLM: 424/190.100
NCLS: 435/069.300; 435/070.210; 435/252.300; 435/320.100;
530/350.000; 536/023.700

L67 ANSWER 2 OF 6 USPATFULL on STN

ACCESSION NUMBER: 2003:258639 USPATFULL

TITLE: 207 human secreted proteins

INVENTOR(S): Ni, Jian, Germantown, MD, UNITED STATES
Ebner, Reinhard, Gaithersburg, MD, UNITED STATES
LaFleur, David W., Washington, DC, UNITED STATES
Moore, Paul A., Germantown, MD, UNITED STATES
Olsen, Henrik S., Gaithersburg, MD, UNITED STATES
Rosen, Craig A., Laytonsville, MD, UNITED STATES
Ruben, Steven M., Olney, MD, UNITED STATES
Soppet, Daniel R., Centreville, VA, UNITED STATES
Young, Paul E., Gaithersburg, MD, UNITED STATES
Shi, Yanggu, Gaithersburg, MD, UNITED STATES
Florence, Kimberly A., Rockville, MD, UNITED STATES
Wei, Ying-Fei, Berkeley, CA, UNITED STATES
Florence, Charles, Rockville, MD, UNITED STATES
Hu, Jing-Shan, Mountain View, CA, UNITED STATES
Li, Yi, Sunnyvale, CA, UNITED STATES
Kyaw, Hla, Frederick, MD, UNITED STATES
Fischer, Carrie L., Burke, VA, UNITED STATES
Ferrie, Ann M., Painted Post, NY, UNITED STATES
Fan, Ping, Potomac, MD, UNITED STATES
Feng, Ping, Gaithersburg, MD, UNITED STATES
Endress, Gregory A., Florence, MA, UNITED STATES
Dillon, Patrick J., Carlsbad, CA, UNITED STATES
Carter, Kenneth C., North Potomac, MD, UNITED STATES
Brewer, Laurie A., St. Paul, MN, UNITED STATES
Yu, Guo-Liang, Berkeley, CA, UNITED STATES
Zeng, Zhizhen, Lansdale, PA, UNITED STATES
Greene, John M., Gaithersburg, MD, UNITED STATES

	NUMBER	KIND	DATE
PATENT INFORMATION:	US 2003181692	A1	20030925
APPLICATION INFO.:	US 2001-933767	A1	20010822 (9)
RELATED APPLN. INFO.:	Continuation-in-part of Ser. No. WO 2001-US5614, filed on 21 Feb 2001, PENDING Continuation-in-part of Ser. No. US 1998-205258, filed on 4 Dec 1998, PENDING		

	NUMBER	DATE
PRIORITY INFORMATION:	US 2000-184836P	20000224 (60)
	US 2000-193170P	20000329 (60)
	US 1997-48885P	19970606 (60)
	US 1997-49375P	19970606 (60)
	US 1997-48881P	19970606 (60)
	US 1997-48880P	19970606 (60)
	US 1997-48896P	19970606 (60)
	US 1997-49020P	19970606 (60)
	US 1997-48876P	19970606 (60)
	US 1997-48895P	19970606 (60)

Searcher : Shears 308-4994

10/081935

US 1997-48884P	19970606 (60)
US 1997-48894P	19970606 (60)
US 1997-48971P	19970606 (60)
US 1997-48964P	19970606 (60)
US 1997-48882P	19970606 (60)
US 1997-48899P	19970606 (60)
US 1997-48893P	19970606 (60)
US 1997-48900P	19970606 (60)
US 1997-48901P	19970606 (60)
US 1997-48892P	19970606 (60)
US 1997-48915P	19970606 (60)
US 1997-49019P	19970606 (60)
US 1997-48970P	19970606 (60)
US 1997-48972P	19970606 (60)
US 1997-48916P	19970606 (60)
US 1997-49373P	19970606 (60)
US 1997-48875P	19970606 (60)
US 1997-49374P	19970606 (60)
US 1997-48917P	19970606 (60)
US 1997-48949P	19970606 (60)
US 1997-48974P	19970606 (60)
US 1997-48883P	19970606 (60)
US 1997-48897P	19970606 (60)
US 1997-48898P	19970606 (60)
US 1997-48962P	19970606 (60)
US 1997-48963P	19970606 (60)
US 1997-48877P	19970606 (60)
US 1997-48878P	19970606 (60)
US 1997-57645P	19970905 (60)
US 1997-57642P	19970905 (60)
US 1997-57668P	19970905 (60)
US 1997-57635P	19970905 (60)
US 1997-57627P	19970905 (60)
US 1997-57667P	19970905 (60)
US 1997-57666P	19970905 (60)
US 1997-57764P	19970905 (60)
US 1997-57643P	19970905 (60)
US 1997-57769P	19970905 (60)
US 1997-57763P	19970905 (60)
US 1997-57650P	19970905 (60)
US 1997-57584P	19970905 (60)
US 1997-57647P	19970905 (60)
US 1997-57661P	19970905 (60)
US 1997-57662P	19970905 (60)
US 1997-57646P	19970905 (60)
US 1997-57654P	19970905 (60)
US 1997-57651P	19970905 (60)
US 1997-57644P	19970905 (60)
US 1997-57765P	19970905 (60)
US 1997-57762P	19970905 (60)
US 1997-57775P	19970905 (60)
US 1997-57648P	19970905 (60)
US 1997-57774P	19970905 (60)
US 1997-57649P	19970905 (60)
US 1997-57770P	19970905 (60)
US 1997-57771P	19970905 (60)
US 1997-57761P	19970905 (60)
US 1997-57760P	19970905 (60)

10/081935

US 1997-57776P 19970905 (60)
US 1997-57778P 19970905 (60)
US 1997-57629P 19970905 (60)
US 1997-57628P 19970905 (60)
US 1997-57777P 19970905 (60)
US 1997-57634P 19970905 (60)
US 1997-70923P 19971218 (60)
US 1998-92921P 19980715 (60)
US 1998-94657P 19980730 (60)
US 1997-70923P 19971218 (60)
US 1998-92921P 19980715 (60)
US 1998-94657P 19980730 (60)

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: HUMAN GENOME SCIENCES INC, 9410 KEY WEST AVENUE,
ROCKVILLE, MD, 20850
NUMBER OF CLAIMS: 23
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 10 Drawing Page(s)
LINE COUNT: 32746

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB The present invention relates to novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating diseases, disorders, and/or conditions related to these novel human secreted proteins.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

INCL INCLM: 536/023.100
INCLS: 530/350.000; 435/325.000; 435/183.000; 435/069.100;
435/320.100
NCL NCLM: 536/023.100
NCLS: 530/350.000; 435/325.000; 435/183.000; 435/069.100;
435/320.100

L67 ANSWER 3 OF 6 USPATFULL on STN

ACCESSION NUMBER: 2003:127198 USPATFULL
TITLE: Death associated kinase containing ankyr in repeats (DAKAR) and methods of use
INVENTOR(S): Bird, Timothy A., Bainbridge Island, WA, UNITED STATES
Holland, Pamela M., Seattle, WA, UNITED STATES
Peschon, Jacques J., Seattle, WA, UNITED STATES
Virca, George D., Bellevue, WA, UNITED STATES

	NUMBER	KIND	DATE
PATENT INFORMATION:	US 2003087411	A1	20030508
APPLICATION INFO.:	US 2002-164080	A1	20020604 (10)

	NUMBER	DATE
PRIORITY INFORMATION:	US 2001-295959P	20010604 (60)
	US 2001-334362P	20011129 (60)

DOCUMENT TYPE: Utility

10/081935

FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: IMMUNEX CORPORATION, LAW DEPARTMENT, 51
UNIVERSITY STREET, SEATTLE, WA, 98101
NUMBER OF CLAIMS: 16
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 4 Drawing Page(s)
LINE COUNT: 5574

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB This invention relates to DAKAR, a new member of the serine/threonine kinase family, methods of making such polypeptides, and to methods of using them to treat conditions associated with apoptosis and epithelial proliferation and differentiation, as well as methods to identify compounds that alter DAKAR-associated cellular activities.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

INCL INCLM: 435/194.000
INCLS: 435/069.100; 435/320.100; 435/325.000; 536/023.200
NCL NCLM: 435/194.000
NCLS: 435/069.100; 435/320.100; 435/325.000; 536/023.200

L67 ANSWER 4 OF 6 USPATFULL on STN

ACCESSION NUMBER: 2003:38356 USPATFULL
TITLE: 125 human secreted proteins
INVENTOR(S): Rosen, Craig A., Laytonsville, MD, UNITED STATES
Feng, Ping, Gaithersburg, MD, UNITED STATES
Ruben, Steven M., Olney, MD, UNITED STATES
Ebner, Reinhard, Gaithersburg, MD, UNITED STATES
Olsen, Henrik S., Gaithersburg, MD, UNITED STATES
Ni, Jian, Germantown, MD, UNITED STATES
Wei, Ying-Fei, Berkeley, CA, UNITED STATES
Soppet, Daniel R., Centreville, VA, UNITED STATES
Moore, Paul A., Germantown, MD, UNITED STATES
Kyaw, Hla, Frederick, MD, UNITED STATES
LaFleur, David W., Washington, DC, UNITED STATES
Shi, Yanggu, Gaithersburg, MD, UNITED STATES
Janat, Fouad, Westerly, RI, UNITED STATES
Endress, Gregory A., Florence, MA, UNITED STATES
Carter, Kenneth C., North Potomac, MD, UNITED STATES
Birser, Charles E., North Potomac, MD, UNITED STATES

	NUMBER	KIND	DATE
PATENT INFORMATION:	US 2003028003	A1	20030206
APPLICATION INFO.:	US 2001-974879	A1	20011012 (9)
RELATED APPLN. INFO.:	Continuation-in-part of Ser. No. US 2001-818683, filed on 28 Mar 2001, PENDING Continuation of Ser. No. US 1999-305736, filed on 5 May 1999, PENDING Continuation-in-part of Ser. No. WO 1998-US23435, filed on 4 Nov 1998, UNKNOWN		

	NUMBER	DATE
PRIORITY INFORMATION:	US 2000-239893P	20001013 (60)
	US 1997-64911P	19971107 (60)
	US 1997-64912P	19971107 (60)

Searcher : Shears 308-4994

10/081935

US 1997-64983P 19971107 (60)
US 1997-64900P 19971107 (60)
US 1997-64988P 19971107 (60)
US 1997-64987P 19971107 (60)
US 1997-64908P 19971107 (60)
US 1997-64984P 19971107 (60)
US 1997-64985P 19971107 (60)
US 1997-66094P 19971117 (60)
US 1997-66100P 19971117 (60)
US 1997-66089P 19971117 (60)
US 1997-66095P 19971117 (60)
US 1997-66090P 19971117 (60)

DOCUMENT TYPE: Utility
FILE SEGMENT: APPLICATION
LEGAL REPRESENTATIVE: HUMAN GENOME SCIENCES INC, 9410 KEY WEST AVENUE,
ROCKVILLE, MD, 20850

NUMBER OF CLAIMS: 23
EXEMPLARY CLAIM: 1
NUMBER OF DRAWINGS: 3 Drawing Page(s)
LINE COUNT: 36277

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB The present invention relates to novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating diseases, disorders, and/or conditions related to these novel human secreted proteins.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

INCL INCLM: 536/023.200
INCLS: 435/069.100; 435/325.000; 435/320.100; 435/183.000
NCL NCLM: 536/023.200
NCLS: 435/069.100; 435/325.000; 435/320.100; 435/183.000

L67 ANSWER 5 OF 6 USPATFULL on STN

ACCESSION NUMBER: 2002:198286 USPATFULL
TITLE: rSOWgp58 protein and its use as an antigen
INVENTOR(S): Hung, Chiung-Yu, Toledo, OH, UNITED STATES
Cole, Garry T., Toledo, OH, UNITED STATES
PATENT ASSIGNEE(S): Medical College of Ohio (U.S. corporation)

	NUMBER	KIND	DATE
PATENT INFORMATION:	US 2002106380	A1	20020808
APPLICATION INFO.:	US 2001-850677	A1	20010507 (9)

	NUMBER	DATE
PRIORITY INFORMATION:	US 2000-202754P	20000508 (60)
DOCUMENT TYPE:	Utility	
FILE SEGMENT:	APPLICATION	
LEGAL REPRESENTATIVE:	EMCH, SCHAFFER, SCHAUB & PORCELLO CO, P O BOX 916, ONE SEAGATE SUITE 1980, TOLEDO, OH, 43697	
NUMBER OF CLAIMS:	32	
EXEMPLARY CLAIM:	1	
NUMBER OF DRAWINGS:	6 Drawing Page(s)	

Searcher : Shears 308-4994

10/081935

LINE COUNT: 1476

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB The present invention provides bacterial plasmids and recombinant SOW proteins that are useful as antigens for serodiagnosis of coccidiomycosis.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

INCL INCLM: 424/191.100

INCLS: 435/007.220; 530/395.000

NCL NCLM: 424/191.100

NCLS: 435/007.220; 530/395.000

L67 ANSWER 6 OF 6 USPATFULL on STN

ACCESSION NUMBER: 2001:208993 USPATFULL

TITLE: Candida albicans gene (CSA1) encoding a mycelial surface antigen, and uses thereof

INVENTOR(S): Bourbonnais, Yves, Cap-Rouge, Canada
Deslauriers, Noella, St-Ferreol-les-Neiges, Canada

PATENT ASSIGNEE(S): Universite Laval, Quebec, Canada (non-U.S. corporation)

	NUMBER	KIND	DATE
PATENT INFORMATION:	US 6320033	B1	20011120
APPLICATION INFO.:	US 1999-351200		19990709 (9)

	NUMBER	DATE
PRIORITY INFORMATION:	CA 1998-2237134	19980710
DOCUMENT TYPE:	Utility	
FILE SEGMENT:	GRANTED	
PRIMARY EXAMINER:	Marschel, Ardin H.	
LEGAL REPRESENTATIVE:	Nixon Peabody LLP	
NUMBER OF CLAIMS:	2	
EXEMPLARY CLAIM:	1	
NUMBER OF DRAWINGS:	19 Drawing Figure(s); 11 Drawing Page(s)	
LINE COUNT:	810	

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

AB The present invention relates to a Candida albicans gene (CSA1) encoding a surface protein. The present invention also relates to the surface protein and methods for using the protein or the gene for the detection, prophylaxis or treatment of candidal infection. The protein encoded is a surface antigen of Candida albicans yeast and mycelial forms, respectively. The mycelial surface antigen was shown to be present predominantly in the terminal third of the hyphal structures. CSA1 is a gene coding for a unique surface antigen.

CAS INDEXING IS AVAILABLE FOR THIS PATENT.

INCL INCLM: 536/023.100

INCLS: 435/006.000; 435/254.220; 536/023.740; 536/024.320

NCL NCLM: 536/023.100

NCLS: 435/006.000; 435/254.220; 536/023.740; 536/024.320

(FILE 'HCAPLUS, MEDLINE, BIOSIS, EMBASE, WPIDS, CONFSCI, SCISEARCH, JICST-EPLUS, JAPIO, USPATFULL' ENTERED AT 09:47:30 ON 04 DEC 2003)

L68 10292 S "COX R"?/AU

Searcher : Shears 308-4994

10/081935

L69 1193 S "MAGEE D"?/AU
L70 5719 S "JIANG C"?/AU
L71 21 S L68 AND L69 AND L70
L72 97 S L68 AND (L69 OR L70)
L73 21 S L69 AND L70

-Author(s)

L75 93 SEA ABB=ON PLU=ON (L71 OR L72 OR L73 OR L68 OR L69 OR
L70) AND (L53 OR L30)
L76 25 DUP REM L75 (68 DUPLICATES REMOVED)

L76 ANSWER 1 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 1

ACCESSION NUMBER: 2003:790176 HCAPLUS

TITLE: Identification of a protective antigen of
Coccidioides immitis by expression
library immunization

AUTHOR(S): Ivey, F. Douglas; **Magee, D. Mitchell**;
Woitaske, Melanie D.; Johnston, Stephen Albert;
Cox, Rebecca A.

CORPORATE SOURCE: Department of Microbiology and Immunology,
University of Texas Health Science Center at San
Antonio, San Antonio, TX, 78229, USA

SOURCE: Vaccine (2003), 21(27-30), 4359-4367
CODEN: VACCDE; ISSN: 0264-410X

PUBLISHER: Elsevier Science Ltd.

DOCUMENT TYPE: Journal

LANGUAGE: English

AB **Coccidioides immitis** is a fungal pathogen of humans and is
classified as a Select Agent. We have identified a new potential
vaccine candidate for this pathogen using cDNA expression library
immunization (ELI). A **C. immitis** spherule-phase
cDNA library containing 800-1000 genes was divided into 10 pools and
each was tested for its protective capacity in BALB/c mice against
i.p. challenge with 2500 arthroconidia of this dimorphic fungus.
The most protective pool, designated Pool 7, was fractionated into
five sublibraries, each containing 60 genes, and of these, only Pool 7-3
induced a significant level of protection in mice. Fractionation of
Pool 7-3 into six sublibraries, each with 10 genes, yielded a
protective fraction, designated Pool 7-3-5. Subsequent fraction of
the latter pool into 10 sublibraries, each with one clone, yielded a
clone designated 7-3-5-5 that was highly protective. Clone 7-3-5-5
was sequenced and found to contain a 672 bp ORF encoding a 224 amino
acid **protein** having a 19 amino acid signal sequence on the
N-terminus and a 15 amino acid C-terminal GPI anchor site. The
7-3-5-5 clone, designated ELI-Antigen 1 (ELI-Ag1), showed partial
homol. with a hypothetical **protein** from *Neurospora crassa*.
This is the first study to identify a protective antigen from a
fungus using ELI, and it is also the first report in which
sequential fractionation of an expression library successfully
identified a single protective gene.

L76 ANSWER 2 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 2

ACCESSION NUMBER: 2003:721561 HCAPLUS

DOCUMENT NUMBER: 139:346526

TITLE: Transfection of murine dendritic cell line (JAWS
II) by a nonviral transfection reagent

AUTHOR(S): Awasthi, Shanjana; **Cox, Rebecca A.**

CORPORATE SOURCE: Dept. of Pathology, University of Texas Health

10/081935

Science Center at San Antonio, San Antonio, TX,
USA
SOURCE: BioTechniques (2003), 35(3), 600-602,604
CODEN: BTNQDO; ISSN: 0736-6205
PUBLISHER: Eaton Publishing Co.
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Dendritic cells are the most potent antigen-presenting cells that initiate and modulate the host immune system. Based on their immunostimulatory activity, a variety of strategies have been developed to use dendritic cells as vaccines and immunotherapeutic agents against infection and cancer. Genetically modified dendritic cells are useful for immunotherapeutic purposes because of their sustained activity in vivo. However, transfection of dendritic cells with plasmid DNA has been very difficult. While the viral transfection is associated with nonspecific activation of dendritic cells, commonly used nonviral transfection reagents have a low efficiency of transfection. Here we describe an improved, simple, less time-consuming transfection protocol using the nonviral nonliposomal lipid polymer, TransIT-TKO transfection reagent, for transfecting murine dendritic cells (JAWS II) with the gene that encodes **Coccidioides immitis** antigen 2 (Ag2). The JAWS II cells were cotransfected with pHYG-enhanced green fluorescent **protein** (EGFP) and pVR1012-**C immitis** Ag2 plasmid DNAs using TransIT-TKO reagent. We reproducibly obtained 30%-50% transfection efficiency. The transfected cells maintained their immature phenotype and were functionally active. In addition, the flexibility of this agent for expressing multiple antigens (GFP and **C. immitis** Ag2) offers an advantage of delivering multiple immunogens.

REFERENCE COUNT: 14 THERE ARE 14 CITED REFERENCES AVAILABLE
FOR THIS RECORD. ALL CITATIONS AVAILABLE
IN THE RE FORMAT

L76 ANSWER 3 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 3

ACCESSION NUMBER: 2002:489694 HCAPLUS

DOCUMENT NUMBER: 137:92359

TITLE: Role of signal sequence in vaccine-induced
protection against experimental
coccidioidomycosis

AUTHOR(S): Jiang, Chengyong; Magee, D.
Mitchell; Ivey, F. Douglas; Cox,
Rebecca A.

CORPORATE SOURCE: Department of Microbiology, University of Texas
Health Science Center at San Antonio, San
Antonio, TX, 78229, USA

SOURCE: Infection and Immunity (2002), 70(7), 3539-3545
CODEN: INFIBR; ISSN: 0019-9567

PUBLISHER: American Society for Microbiology

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The vaccine efficacy of the gene sequence encoding the signal **peptide** of the antigen known as antigen 2 or proline-rich antigen (Ag2/PRA), an immunodominant antigen present in the cell wall of the fungal pathogen **Coccidioides immitis**, was investigated in a murine model of **coccidioidomycosis**. Expression plasmids for Ag2/PRA(1-18) DNA (signal sequence), Ag2/PRA(19-194) DNA (lacking the signal sequence), and

10/081935

Ag2/PRA(1-194) DNA (full length) were inserted in the pVR1012 vector, and the constructs were used to vaccinate the highly susceptible BALB/c mouse strain. Immunization with the signal gene sequence significantly reduced the fungal burden in the lungs and spleens of mice 12 days after i.p. challenge with a LD of 2500 *C. immitis* arthroconidia, to a level comparable to the protection induced in mice immunized with the full-length Ag2/PRA(1-194) DNA. The Ag2/PRA(19-194) gene protected mice but to a significantly lower level than the signal sequence or the full-length Ag2 gene. The immunizing capacity of Ag2/PRA(1-18) was not attributable to a nonspecific immunostimulatory effect of DNA, as evidenced by the fact that mice immunized with a frameshift mutation of Ag2/PRA(1-18) were not protected against challenge. Furthermore, a synthetic **peptide** corresponding to the translated sequence of Ag2/PRA(1-18) DNA protected mice, albeit at a lower level than the Ag2/PRA(1-18) DNA vaccine. The protection induced with the signal gene vaccine correlated with the production of gamma interferon when splenocytes from Ag2/PRA(1-18)-immunized mice were stimulated with recombinant full-length Ag2 and was not associated with the production of anti-*Coccidioides* IgG antibody. This is the first study to establish that a signal **peptide** sequence alone, administered as a gene vaccine or synthetic **peptide**, can induce protective immunity against a microbial pathogen.

REFERENCE COUNT: 36 THERE ARE 36 CITED REFERENCES AVAILABLE
FOR THIS RECORD. ALL CITATIONS AVAILABLE
IN THE RE FORMAT

L76 ANSWER 4 OF 25 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on
STN

ACCESSION NUMBER: 2002:597076 BIOSIS

DOCUMENT NUMBER: PREV200200597076

TITLE: Induction of gene expression in *Coccidioides*
immitis during growth inside murine macrophages.

AUTHOR(S): **Magee, D. M.** [Reprint author]; **Woitaske, M.**
D. [Reprint author]; **Wickes, B. L.** [Reprint author];
Cox, R. A. [Reprint author]

CORPORATE SOURCE: University of Texas Health Science Center at San
Antonio, San Antonio, TX, USA

SOURCE: Abstracts of the General Meeting of the American
Society for Microbiology, (2002) Vol. 102, pp. 205.
print.
Meeting Info.: 102nd General Meeting of the American
Society for Microbiology. Salt Lake City, UT, USA.
May 19-23, 2002. American Society for Microbiology.
ISSN: 1060-2011.

DOCUMENT TYPE: Conference; (Meeting)
Conference; Abstract; (Meeting Abstract)

LANGUAGE: English

ENTRY DATE: Entered STN: 20 Nov 2002

Last Updated on STN: 20 Nov 2002

AB *Coccidioides immitis* is a dimorphic fungus of the American
southwest growing in the soil as mycelia and within the host as the
spherule/endospore parasitic phase. We hypothesized that genes that
are upregulated when *C. immitis* grows in vivo
might be associated with virulence of this primary fungal pathogen.
To identify potential virulence-associated genes, we utilized
differential display, reverse transcriptase, polymerase chain

10/081935

reaction (DD-RT-PCR) to assess induction of gene expression within *C. immitis* spherules during intracellular parasitism of murine macrophages. Macrophage monolayers were infected with *C. immitis* arthroconidia and, after approximately 36 hours of culture, RNA was obtained from the developing macrophage-derived spherules and compared to RNA from spherules grown in tissue culture medium alone. Differential display analyses identified 11 gene fragments specific for the macrophage-derived spherules. These products were cloned into a TA vector, sequenced, and analyzed for homology with the Genbank databases. One clone had significant homology to a *Magnaporthe grisea* appressorium stage cDNA library gene and *Saccharomyces cerevisiae* glycogen phosphorylase. The complete cDNA clone was recovered using 5' and 3' RACE and using this as a probe, we obtained and sequenced a genomic clone for *C. immitis* glycogen phosphorylase. The gene had seven predicted introns and a translation of the coding region indicated that the **protein** would contain 881 amino acids with a molecular weight of 100.6 kDa and a pI of 5.75. Northern analysis confirmed that glycogen phosphorylase message was differentially expressed by spherules in parasitized macrophages. Differential display may provide a tool to delineate virulence-associated genes of *C. immitis* or help define metabolic targets for rationally designed antifungal drugs.

L76 ANSWER 5 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 4

ACCESSION NUMBER: 2000:208899 HCAPLUS

DOCUMENT NUMBER: 132:247984

TITLE: The x-ray structure of a chitinase from the

AUTHOR(S): pathogenic fungus *Coccidioides immitis*
Hollis, Thomas; Monzingo, Arthur F.; Bortone,
Kara; Ernst, Stephen; Cox, Rebecca;
Robertus, Jon D.

CORPORATE SOURCE: Institute of Cellular and Molecular Biology,
Department of Chemistry and Biochemistry,
University of Texas, Austin, TX, 78712, USA

SOURCE: Protein Science (2000), 9(3), 544-551
CODEN: PRCLIE; ISSN: 0961-8368

PUBLISHER: Cambridge University Press

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The x-ray crystal structure of chitinase from *C.*

immitis was solved to 2.2 Å resolution. Like other members of the class 18 hydrolase family, this 427-residue **protein** was an 8-stranded β/α -barrel. Although lacking an N-terminal chitin-anchoring domain, the enzyme closely resembled chitinase from *Serratia marcescens*. Among the conserved features were 3 cis **peptide** bonds, all involving conserved active site residues. The active site was formed from conserved residues such as Trp-47, Trp-131, Trp-315, Trp-378, Tyr-239, Tyr-293, Arg-52 and Arg-295. Glu-171 was the catalytic acid in the hydrolytic mechanism; when it was mutated to Gln, enzyme activity was abolished. Allosamidin is a substrate analog that strongly inhibits the class 18 enzymes. Its binding to the chitinase, hevine, was previously observed and conserved structural features of the 2 enzymes were used to predict the inhibitor's binding to the fungal enzyme.

REFERENCE COUNT: 39 THERE ARE 39 CITED REFERENCES AVAILABLE
FOR THIS RECORD. ALL CITATIONS AVAILABLE

10/081935

IN THE RE FORMAT

L76 ANSWER 6 OF 25 EMBASE COPYRIGHT 2003 ELSEVIER INC. ALL RIGHTS
RESERVED. on STN

ACCESSION NUMBER: 1999185649 EMBASE

TITLE: Construction of a single-chain interleukin-12-
expressing retroviral vector and its application in
cytokine gene therapy against experimental
coccidioidomycosis.

AUTHOR: Jiang C.; Magee D.M.; Cox
R.A.

CORPORATE SOURCE: C. Jiang, Dept. of Clinical Investigation, Texas
Center for Infectious Disease, 2303 SE Military Dr.,
San Antonio, TX 78223, United States.
chengyong.jiang@tdh.state.tx.us

SOURCE: Infection and Immunity, (1999) 67/6 (2996-3001).
Refs: 45

ISSN: 0019-9567 CODEN: INFIBR

COUNTRY: United States

DOCUMENT TYPE: Journal; Article

FILE SEGMENT: 004 Microbiology

026 Immunology, Serology and Transplantation

LANGUAGE: English

SUMMARY LANGUAGE: English

AB T-cell-mediated immunity is an important determinant in protection
against primary infection with **Coccidioides immitis**, a
dimorphic fungal pathogen that causes the disease
coccidioidomycosis. To determine if interleukin-12 (IL-12)
gene therapy could potentiate host response against **C.**
immitis, we constructed a single-chain cDNA encoding the p40
and p35 subunits linked by a polylinker and, using a retroviral
vector, transfected J774 macrophages with the construct. The
transduced J774 cells expressed IL-12 in vitro, with a mean
concentration of 28,440 pg from 106 cells in 48 h as measured by an
IL-12 (p75)-specific enzyme-linked immunosorbent assay. The secreted
IL-12 was biologically active, as judged by its ability to induce
the production of gamma interferon (IFN- γ) by spleen cells
from BALB/c mice. Treatment of the highly susceptible BALB/c mouse
strain with the IL-12- transduced J774 cells inhibited **C.**
immitis growth in tissues from mice challenged by a
pulmonary route, as evidenced by 1.37-, 2.59-, and 1.22-log
reductions in the number of CFU in the lungs, spleens, and livers,
respectively, compared to the fungal load in mice given
vector-transduced J774 cells. The protective effect of IL-12 gene
therapy was accompanied by increased levels of IFN- γ in the
lungs and sera of mice treated with IL-12- transduced J774 cells and
the constitutive production of IFN- γ by their spleen cells
cultured in vitro. These results suggest that IL-12 gene therapy
could be used as adjunct therapy for **coccidioidomycosis**.

L76 ANSWER 7 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 5

ACCESSION NUMBER: 1999:83280 HCAPLUS

DOCUMENT NUMBER: 130:266091

TITLE: Genetic vaccination against **Coccidioides**
immitis: comparison of vaccine efficacy of
recombinant antigen 2 and antigen 2 cDNA

AUTHOR(S): Jiang, Chengyong; Magee, D.
Mitchell; Quitugua, Teresa N.; Cox,

10/081935

Rebecca A.
CORPORATE SOURCE: Department of Clinical Investigation, Texas
Center for Infectious Disease, San Antonio, TX,
78223, USA
SOURCE: Infection and Immunity (1999), 67(2), 630-635
CODEN: INFIBR; ISSN: 0019-9567
PUBLISHER: American Society for Microbiology
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Previous studies from our laboratory established that C-ASWS, an
alkali-soluble, water-soluble extract from cell walls of **Coccidioides**
immitis, protects mice against lethal challenge with this fungus.
The C-ASWS extract contains a glycosylated **protein**,
designated antigen 2 (Ag2), and a polysaccharide antigen. We
recently cloned Ag2 cDNA and showed that the recombinant fusion
protein elicited strong delayed-type hypersensitivity
responses in immunized mice. This investigation was undertaken to
determine if the recombinant Ag2 **protein**, expressed as an
Ag2-glutathione S-transferase (GST) fusion **protein**, or Ag2
cDNA would protect mice against lethal challenge with **C.**
immitis. The recombinant Ag2-GST **protein**
protected BALB/c mice against i.p. challenge with 250 arthroconidia,
as assessed by a decrease in fungal CFU in tissues. The
Ag2-GST-immunized mice did not show, however, an increased survival
during a 30-day period postinfection. By contrast, immunization of
mice with Ag2 cDNA ligated into the pVR1012 plasmid engendered
protection against i.p. challenge with 2,500 arthroconidia and
against pulmonary challenge with 50 arthroconidia. Vaccine efficacy
paralleled the development of delayed-type hypersensitivity
responses to **C. immitis** antigen. Whereas mice
vaccinated with the recombinant Ag2-GST **protein** did not
mount footpad hypersensitivity to C-ASWS or the recombinant Ag2-GST
protein, mice vaccinated with the pVR1012-Ag2 construct
mounted a strong footpad hypersensitivity and their spleen cells
secreted gamma interferon upon in vitro stimulation with the
Ag2-containing C-ASWS extract. This is the first investigation to show that
genetic immunization can protect against lethal challenge with
C. immitis.
REFERENCE COUNT: 50 THERE ARE 50 CITED REFERENCES AVAILABLE
FOR THIS RECORD. ALL CITATIONS AVAILABLE
IN THE RE FORMAT

L76 ANSWER 8 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 6
ACCESSION NUMBER: 1999:2211 HCAPLUS
DOCUMENT NUMBER: 130:135866
TITLE: Crystallization and preliminary x-ray analysis
of a chitinase from the fungal pathogen
Coccidioides immitis
AUTHOR(S): Hollis, Thomas; Monzingo, Arthur F.; Bortone,
Kara; Schelp, Elisabeth; Cox, Rebecca;
Robertus, Jon D.
CORPORATE SOURCE: Institute of Cellular and Molecular Biology,
Department of Chemistry and Biochemistry,
University of Texas, Austin, TX, 78712, USA
SOURCE: Acta Crystallographica, Section D: Biological
Crystallography (1998), D54(6, Pt. 2), 1412-1413
CODEN: ABCRE6; ISSN: 0907-4449
PUBLISHER: Munksgaard International Publishers Ltd.

10/081935

DOCUMENT TYPE: Journal
LANGUAGE: English

AB Chitinase from *C. immitis* was expressed as a fusion **protein** with glutathione-S-transferase (GST), which aided in its purification. After cleavage from GST, the chitinase was crystallized from 30% PEG 4000 in 0.1M NaOAc, pH 4.6. The crystals had a tetragonal crystal lattice and belonged to space group P4₁2₁2 or P4₃2₁2 and diffracted to 2.2 Å resolution. The unit-cell parameters were a = b = 91.2, c = 95.4 Å; there is only 1 chitinase mol. in the asym. unit.

REFERENCE COUNT: 12 THERE ARE 12 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L76 ANSWER 9 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 7

ACCESSION NUMBER: 1997:655696 HCAPLUS

DOCUMENT NUMBER: 127:330024

TITLE: Mapping of a *Coccidioides immitis*-specific epitope that reacts with complement-fixing antibody

AUTHOR(S): Yang, Michael C.; Magee, D. Mitchell; Cox, Rebecca A.

CORPORATE SOURCE: Department of Clinical Investigation, Texas Center for Infectious Disease, San Antonio, TX, 78223, USA

SOURCE: Infection and Immunity (1997), 65(10), 4068-4074
CODEN: INFIBR; ISSN: 0019-9567

PUBLISHER: American Society for Microbiology

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The authors have previously cloned the cDNA fragment that encodes the complement fixation antigen of *Coccidioides immitis*. The recombinant **protein** was highly sensitive in detecting CF antibody in sera from patients with **coccidioidomycosis** but was not specific to *C. immitis*, as evidence by its reactivity with sera from patients with histoplasmosis and, to lesser extent, blastomycosis. The authors undertook this study to determine if the epitope(s) that reacts with CF antibody is the same or differs from the epitopes that are shared with *Histoplasma capsulatum* and *Blastomyces dermatitidis*. PCR-generated CF/chitinase cDNA fragments were cloned and examined for their reactivity in enzyme-linked immunosorbent assays using sera from patients with **coccidioidomycosis**, histoplasmosis, or blastomycosis. A **peptide** domain comprised of amino acid residues 20 through 310 was shown to express an epitope(s) that is specific to anti-*Coccidioides* CF antibody. The **peptide** detected serum antibody in 21 (95%) of 22 patients with active **coccidioidomycosis** and was without reactivity with sera from 20 patients with histoplasmosis, 15 patients with blastomycosis, and 14 healthy subjects. Antibody titers to the recombinant **peptide** directly correlated with CF antibody titers ($P < 0.01$), and preadsorption of reference CF antiserum with the **peptide** ablated the reactivity of the antiserum in the immunodiffusion assay for CF antibody. The delineation of a recombinant **peptide** that has both sensitivity and specificity will provide a valuable tool for detecting CF antibody and for evaluating the role of CF antibody in the host response to *C. immitis*.

10/081935

L76 ANSWER 10 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 8
ACCESSION NUMBER: 1997:482804 HCAPLUS
DOCUMENT NUMBER: 127:204149
TITLE: Identification of a **Coccidioides**
immitis antigen 2 domain that expresses
B-cell-reactive epitopes
AUTHOR(S): Zhu, Yufan; Tryon, Vic; **Magee, D.**
Mitchell; Cox, Rebecca A.
CORPORATE SOURCE: Department of Clinical Investigation, Texas
Center for Infectious Disease, San Antonio, TX,
78223, USA
SOURCE: Infection and Immunity (1997), 65(8), 3376-3380
CODEN: INFIBR; ISSN: 0019-9567
PUBLISHER: American Society for Microbiology
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Antigen 2 (Ag2), a major immunoreactive component of **C.**
immitis mycelium- and spherule-phase cell walls, was
recently cloned and was shown to elicit T-cell responses in
Coccidioides-immune mice. Here, the authors evaluated
recombinant Ag2 (rAg2) and PCR-generated Ag2 truncations for
expression of B-cell-reactive epitopes in ELISA and immunoblot
assays with sera from patients with active
coccidioidomycosis, a hyperimmune goat anti-Ag2 serum, and a
murine anti-Ag2 monoclonal antibody that recognizes a conformational
epitope. The results established that rAg2 expresses both linear
and conformational B-cell-reactive epitopes which are localized to a
domain comprised of amino acids 19-96 (designated A19-96).
Truncations designed to identify epitopes within the A19-96 domain
yielded fragments that either were nonreactive (A62-194, A19-61, and
A49-79) or showed reduced reactivity (A19-79). Hence, A19-96 was
the shortest domain expressing epitopes recognized by the panel of
antibodies. The prevalence of antibodies to the A19-96 domain was
evaluated in ELISAs of sera from 28 **coccidioidomycosis**
patients. Antibody reactivity was detected in 79% of the patients'
sera, and the level of antibody reactivity was directly correlated
with disease severity. Whereas patients with pulmonary disease
showed a mean response (A405) of 0.16, patients with disseminated
coccidioidomycosis showed a mean response of 0.69. No
reactivity was detected with sera from histoplasmosis or
blastomycosis patients. The production of a recombinant **peptide**
that expresses **C. immitis**-specific Ag2 epitopes
provides a useful reagent for examining the role of anti-Ag2 antibodies
in **coccidioidomycosis**.

L76 ANSWER 11 OF 25 MEDLINE on STN DUPLICATE 9
ACCESSION NUMBER: 97160980 MEDLINE
DOCUMENT NUMBER: 97160980 PubMed ID: 9008276
TITLE: Recombinant **Coccidioides immitis**
complement-fixing antigen: detection of an epitope
shared by **C. immitis**, Histoplasma
capsulatum, and Blastomyces dermatitidis.
AUTHOR: Yang M C; **Magee D M**; Kaufman L; Zhu Y;
Cox R A
CORPORATE SOURCE: Department of Clinical Investigation, Texas Center
for Infectious Disease, San Antonio 78223, USA.
CONTRACT NUMBER: AI12431 (NIAID)

Searcher : Shears 308-4994

10/081935

SOURCE: CLINICAL AND DIAGNOSTIC LABORATORY IMMUNOLOGY, (1997
Jan) 4 (1) 19-22.
Journal code: 9421292. ISSN: 1071-412X.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 199704
ENTRY DATE: Entered STN: 19970422
Last Updated on STN: 19970422
Entered Medline: 19970408

AB We undertook an investigation to assess the utility of a recombinant **Coccidioides** immitis complement-fixing (CF) antigen for detecting CF antibody in sera from patients with **coccidioidomycosis**. Enzyme-linked immunosorbent assays established that recombinant CF antigen and, for comparison, a commercially available **coccidioidin** were reactive with 19 of 19 sera from patients with active **coccidioidomycosis**. The recombinant antigen was significantly more sensitive than **coccidioidin**. The median titer obtained when patients' sera were assayed against recombinant CF antigen was 1:51,200 compared to 1:25,600 with **coccidioidin** ($P < 0.027$). The recombinant antigen was also more effective in distinguishing the antibody levels in sera from patients with pulmonary **coccidioidomycosis** than in sera from those with disseminated disease. Whereas patients with pulmonary disease showed a median antibody titer of 1:25,600, those with multifocal disease showed a median titer of 1:102,400 ($P < 0.028$). The recombinant CF antigen was found, however, to express an epitope(s) that reacted with sera from 6 of 12 patients with histoplasmosis and 2 of 12 patients with blastomycosis.

L76 ANSWER 12 OF 25 MEDLINE on STN
ACCESSION NUMBER: 96355868 MEDLINE
DOCUMENT NUMBER: 96355868 PubMed ID: 8751906
TITLE: Interleukin-12 regulation of host defenses against **Coccidioides** immitis.
AUTHOR: Magee D M; Cox R A
CORPORATE SOURCE: Department of Clinical Investigation, Texas Center for Infectious Diseases, San Antonio 78223, USA..
mmagee@tcid.tdh.state.tx.us
CONTRACT NUMBER: AI23555 (NIAID)
SOURCE: INFECTION AND IMMUNITY, (1996 Sep) 64 (9) 3609-13.
Journal code: 0246127. ISSN: 0019-9567.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals; AIDS
ENTRY MONTH: 199610
ENTRY DATE: Entered STN: 19961015
Last Updated on STN: 19961015
Entered Medline: 19961003

AB We have previously reported on the alternate regulation of gamma interferon (IFN-gamma) and interleukin-4 (IL-4) in inbred mouse strains which differ in their susceptibility to **Coccidioides** immitis. The genetically resistant DBA/2 mice manifest a predominant T-helper 1 (Th1) response, with early production of IFN-gamma, whereas susceptible BALB/c mice show an early production

10/081935

of the Th2 cytokine IL-4. Since IL-12 is one cytokine that can act early during host defenses to promote the differentiation of cytokine production towards IFN-gamma and thus may promote expression of a protective immune response, we investigated the role of IL-12 in resistance to *C. immitis*.

Administration of recombinant IL-12 to the susceptible mouse strain before and after systemic (intraperitoneal) challenge with *C. immitis* significantly ameliorated the course of the disease, as measured by a reduction in the fungal load in the lungs, liver, and spleen. Analysis of the cytokine mRNA in lungs from infected BALB/c mice revealed that the protective effect of recombinant IL-12 was accompanied by a shift from a Th2 to a Th1 response. The importance of IL-12 in resistance to this fungus was further established by showing that neutralization of endogenous IL-12 in the resistant DBA/2 mouse strain led to a significant increase in the fungal burden in pulmonary and extrapulmonary tissues. These results establish that IL-12 plays a pivotal role in the host defense against systemic challenge with *C. immitis*.

L76 ANSWER 13 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 10
ACCESSION NUMBER: 1996:391996 HCAPLUS
DOCUMENT NUMBER: 125:55750
TITLE: Molecular cloning and characterization of
Coccidioides immitis antigen 2 cDNA
AUTHOR(S): Zhu, Yufan; Yang, Chunmu; Magee, D.
Mitchell; Cox, Rebecca A.
CORPORATE SOURCE: Dep. Clinical Investigation, Texas Center
Infectious Disease, San Antonio, TX, 78223, USA
SOURCE: Infection and Immunity (1996), 64(7), 2695-2699
CODEN: INFIBR; ISSN: 0019-9567
PUBLISHER: American Society for Microbiology
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Previous expts. have provided evidence that *Coccidioides immitis* antigen 2 (Ag2) is a major T-cell-reactive component of mycelia and spherule cell walls. Here the authors report the identification and cloning of the cDNA that encodes Ag2 from a lambda ZAP cDNA expression library constructed from spherule-derived RNA. DNA sequence anal. established that the 1,255-bp clone contains a 174-bp 5' untranslated region, a 582-bp open reading frame which encodes for a **protein** consisting of 194 amino acids, and a 375-bp 3' untranslated region, including a poly(A) tail. The recombinant Ag2 **protein** has a predicted mol. mass of 19.5 kDa and contains an 18-amino-acid N terminus which has been tentatively identified as a signal **peptide**. The Ag2 cDNA was ligated into the pGEX-4T-3 vector and expressed in *Escherichia coli* TG-1 cells as a glutathione S-transferase fusion **protein**. The recombinant fusion **protein** showed reactivity with sera from patients with **coccidioidomycosis** and elicited delayed-type footpad hypersensitivity responses in *Coccidioides*-immune mice. These results suggest that the Ag2 cDNA can be used for the large-scale production of this immunol. important **protein**.

L76 ANSWER 14 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 11
ACCESSION NUMBER: 1996:327965 HCAPLUS
DOCUMENT NUMBER: 125:28509

10/081935

TITLE: Molecular cloning and characterization of the **Coccidioides immitis** complement fixation/chitinase antigen

AUTHOR(S): Yang, Chunmu; Zhu, Yufan; Magee, D. Mitchell; Cox, Rebecca A.

CORPORATE SOURCE: Department Clinical Investigation, Texas Center Infectious disease, San Antonio, TX, 78223, USA

SOURCE: Infection and Immunity (1996), 64(6), 1992-1997
CODEN: INFIBR; ISSN: 0019-9567

PUBLISHER: American Society for Microbiology

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Detection of anti-**Coccidioides** complement-fixing (CF) antibody is a valuable diagnostic and prognostic aid in **coccidioidomycosis**. The CF antibody response is directed against a heat-labile antigen that has chitinase activity, hereafter referred to as the CF/chitinase **protein**. To identify and clone this immunoreactive enzyme, the authors constructed a **Coccidioides immitis** cDNA lambda ZAP expression library from spherule RNA and detected fusion **peptides** expressing CF epitopes by immunoscreening. A cDNA clone consisting of 1623 bp was identified, sequenced, and found to contain a single open reading frame that encodes a **protein** of 47 kDa with 427 amino acids. Deduced amino acid sequence analyses showed that the cloned CF/chitinase cDNA contains a 35-amino-acid region, beginning at Ser-18 and ending at Arg-52, which has 92% homol. with the reported N-terminal amino acid sequence of authentic CF/chitinase **protein**. The first 17 amino acids in the deduced sequence of the cloned cDNA are not present on the mature CF/chitinase **protein**, suggesting that it may be a signal **peptide**. Expression of the CF/chitinase cDNA insert by using the pGEX-4T-3 vector yields a fusion **peptide** that bears CF-specific epitopes and shows chitinase activity. The CF/chitinase clone will enable large-scale production of the recombinant CF antigen for use in immunoassays and facilitate studies on the role of chitinase in the morphogenesis of **C immitis**.

L76 ANSWER 15 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 12

ACCESSION NUMBER: 1996:737195 HCAPLUS

DOCUMENT NUMBER: 126:86206

TITLE: **Coccidioides immitis** Antigen 2: analysis of gene and **protein**

AUTHOR(S): Zhu, Yufan; Yang, Chunmu; Magee, D. Mitchell; Cox, Rebecca A.

CORPORATE SOURCE: Department of Clinical Investigation, Texas Center for Infectious Disease, 2303 SE Military Drive, San Antonio, TX, 78223, USA

SOURCE: Gene (1996), 181(1/2), 121-125
CODEN: GENED6; ISSN: 0378-1119

PUBLISHER: Elsevier

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Antigen 2 is a glycosylated **protein** present in the cell walls of the dimorphic fungus **Coccidioides immitis**. Using oligodeoxyribonucleotide (oligo) primers based on the sequences of Ag2 cDNA, the gene encoding Ag2 was cloned from genomic DNA derived from the mycelial phase of **C. immitis** by PCR. Nucleotide (nt) sequence analyses showed a 582 base pair (bp) ORF

10/081935

disrupted by two introns which are 78bp and 101bp long. The deduced primary translation product consists of 194 amino acids (aa), contains an N-terminal putative signal sequence to allow transport into the endoplasmic reticulum, and a C-terminal putative signal sequence to enable a GPI anchor addition. Putative GPI anchor/cleavage site and O-glycosylation sites, as well as phosphorylation and myristoylation sites are also present. On the basis of these analyses, the authors predict that a prepro-Ag2 undergoes a post-translational modification to yield the mature glycosylated Ag2 **protein** which is anchored on the extracellular plasma membrane of mycelial and spherule-phase cells.

L76 ANSWER 16 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 13
ACCESSION NUMBER: 1995:781432 HCAPLUS
DOCUMENT NUMBER: 123:196430
TITLE: Roles of gamma interferon and interleukin-4 in
genetically determined resistance to

Coccidioides immitis
AUTHOR(S): Magee, D. Mitchell; Cox, Rebecca
A.

CORPORATE SOURCE: Dep. Res. Immunol., Texas Cent. Infectious
Diseases, San Antonio, TX, 78223, USA
SOURCE: Infection and Immunity (1995), 63(9), 3514-19
CODEN: INFIBR; ISSN: 0019-9567
PUBLISHER: American Society for Microbiology
DOCUMENT TYPE: Journal
LANGUAGE: English

AB The profiles of γ interferon (IFN- γ) and interleukin-4 (IL-4) production were valuated during the course of **coccidioidomycosis** in 2 inbred mouse strains which differ in their susceptibility to **C. immitis**. Cytokine responses, measured at the mol. and **protein** levels, showed increased levels of IFN- γ in lung exts. from mice of the resistant DBA/2 strain after a pulmonary challenge, whereas the susceptible BALB/c strain manifested a predominant IL-4 response. The importance of these cytokines in host defense against **C. immitis** was established by treating the mice with recombinant cytokines or neutralizing anticytokine monoclonal antibodies. Treatment of the susceptible BALB/c mice with recombinant murine IFN- γ protected mice against systemic challenge, and in the reciprocal experiment, the administration of an anti-IFN- γ monoclonal antibody to the resistant DBA/2 mice decreased their capacity to control disease. Although the treatment of DBA/2 mice with recombinant IL-4 did not alter the disease, neutralization of endogenous IL-4 in infected BALB/c mice by administration of a neutralizing anti-IL-4 antibody led to a reduction in the fungal load in their tissues. Apparently, IFN- γ plays a pivotal role in resistance to **C. immitis**, whereas IL-4 down-regulates protective immunity against **C. immitis**.

L76 ANSWER 17 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 14
ACCESSION NUMBER: 1993:426295 HCAPLUS
DOCUMENT NUMBER: 119:26295
TITLE: Production of a murine monoclonal antibody that
recognizes an epitope specific to

Coccidioides immitis antigen 2
AUTHOR(S): Cox, Rebecca A.; Dolan, Matthew J.;

10/081935

CORPORATE SOURCE: Magee, D. Mitchell; Galgiani, John N.
Dep. Res. Immunol., San Antonio State Chest
Hosp., San Antonio, TX, 78223, USA
SOURCE: Infection and Immunity (1993), 61(5), 1895-9
CODEN: INFIBR; ISSN: 0019-9567
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Antigen 2 (Ag2) has been implicated as a T-cell-reactive component of the pathogenic fungus *C. immitis*. This report describes the production of a murine monoclonal antibody (MAb) of the IgG2a isotype that recognizes an epitope specific to *C. immitis* Ag2. This specificity was evidenced by the finding that the MAb did not recognize other antigens present in *coccidioidin* or spherulin and did not show reactivity with antigenic exts. from *Histoplasma capsulatum* or *Blastomyces dermatitidis*. The epitope was labile to enzymic digestion with pronase but resistant to treatment with glycolytic enzymes and to periodate oxidation. This **peptide** epitope appears to require conformational structure because it was not recognized by the MAb in immunoblots of antigen that had been electrophoresed in polyacrylamide gels under denaturing, reducing conditions. Immunoaffinity chromatog. of spherulin on columns containing the MAb established that the MAb was effective as a ligand for isolating Ag2 from heterogeneous exts. The production of a MAb which recognizes an Ag2-specific epitope and its utility as a ligand for isolating Ag2 will provide a valuable reagent for studies of this immunol. important antigen.

L76 ANSWER 18 OF 25 EMBASE COPYRIGHT 2003 ELSEVIER INC. ALL RIGHTS
RESERVED. on STN DUPLICATE 15

ACCESSION NUMBER: 92245388 EMBASE
DOCUMENT NUMBER: 1992245388
TITLE: Localization of the tube precipitin and complement fixation antigens of *Coccidioides immitis* by immunoelectron microscopy with murine monoclonal antibodies.
AUTHOR: Cox R.A.; Sun S.H.; Dolan M.J.; Harrison J.L.
CORPORATE SOURCE: Department of Research Immunology, San Antonio State Chest Hospital, San Antonio, TX 78223, United States
SOURCE: Infection and Immunity, (1992) 60/8 (3315-3324).
ISSN: 0019-9567 CODEN: INFIBR
COUNTRY: United States
DOCUMENT TYPE: Journal; Article
FILE SEGMENT: 004 Microbiology
005 General Pathology and Pathological Anatomy
026 Immunology, Serology and Transplantation
LANGUAGE: English
SUMMARY LANGUAGE: English

AB The cellular localization of the tube precipitin (TP) and complement fixation (CF) antigens of *Coccidioides immitis* was examined by immunoelectron microscopy with murine immunoglobulin G1 monoclonal antibodies directed against the TP and CF antigens, respectively. Immunoelectron microscopic analyses of saprobic- and parasitic-phase cells showed that the TP antigen is present at a high concentration within the inner cell wall layer and along the plasma membrane. The antigen was also detected, at a lesser concentration, within cytoplasmic vacuoles. In contrast to the

10/081935

predominant localization of the TP antigen in the cell walls, the CF antigen resides primarily within the cytoplasm, where it appears to be dispersed throughout the cytoplasm rather than associated with a specific cytoplasmic organelle. A sparse amount of the CF antigen within the inner cell walls was also demonstrable. The localization of the TP and CF antigens throughout the morphogenetic phases of **C. immitis** has important implications in antigen production and in analyses of host response in **coccidioidomycosis**.

L76 ANSWER 19 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 16
ACCESSION NUMBER: 1991:533582 HCAPLUS
DOCUMENT NUMBER: 115:133582
TITLE: An immunoreactive apoglycoprotein purified from **Coccidioides immitis**
AUTHOR(S): Dugger, Kris O.; Galgiani, John N.; Ampel, Neil M.; Sun, Sung H.; **Magee, D. Mitchell**; Harrison, Jeff; Law, John H.
CORPORATE SOURCE: Med. Res. Serv., Veterans Aff. Med. Cent., Tucson, AZ, 85723, USA
SOURCE: Infection and Immunity (1991), 59(7), 2245-51
CODEN: INFIBR; ISSN: 0019-9567
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Deglycosylation of glycoproteins in a lysate of spherules of **C. immitis** has permitted purification and partial characterization of a proline-rich pronase-sensitive antigen. Moreover, soluble antigen specifically stimulated lymphocytes from persons with dermal delayed-type hypersensitivity to **coccidioidal** antigens. When related to reference **coccidioidin** by tandem 2-dimensional immunoelectrophoresis, the antigen fused in the anodal region with a specific reference antigen (antigen 2). It did not show identity with **coccidioidal** antigens used in conventional serol. assays. Although immunoblots of the purified **protein** with monospecific rabbit antiserum showed a single antigen at 33 kDa, the parent spherule lysate bound the same antibody in a broad band between 70 and >200 kDa, which could be explained by microheterogeneity of glycosylation. Immunoelectron microscopy using affinity-purified human antibodies localized the antigen to the cell wall and internal septa of spherules. The apoglycoprotein may be important in human immune responses to **coccidioidal** infection.

L76 ANSWER 20 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 17
ACCESSION NUMBER: 1991:490408 HCAPLUS
DOCUMENT NUMBER: 115:90408
TITLE: Production and characterization of a monoclonal antibody to the complement fixation antigen of **Coccidioides immitis**
AUTHOR(S): Dolan, Matthew J.; **Cox, Rebecca A.**
CORPORATE SOURCE: Dep. Res. Immunol., San Antonio State Chest Hosp., San Antonio, TX, 78223, USA
SOURCE: Infection and Immunity (1991), 59(6), 2175-80
CODEN: INFIBR; ISSN: 0019-9567
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Detection of complement-fixing antibody to **coccidioidin** by using the complement fixation test or an immunodiffusion assay for

10/081935

complement-fixing antibody (IDCF) is widely viewed as the most useful immunodiagnostic test for **coccidioidomycosis**. Here is reported the production of an IgG1 monoclonal antibody (MAB) to the IDCF antigen for use as a biospecific ligand for purifying the IDCF antigen on solid-phase immunosorbents and for use as a reagent for screening genomic or cDNA expression libraries from **C. immitis**. BALB/c mice were immunized by i.m. injections of **coccidioidin** in adjuvant, followed by an intrasplenic booster injection of **coccidioidin** in saline. The spleen cells were fused with SP2/0 Ag14 myeloma cells, and the fusion products were screened for IgG antibody to **coccidioidin** by using an ELISA. Pos. hybridomas were cloned and evaluated for reactivity to the IDCF antigen by 2-dimensional immunoelectrophoresis and by immunoblotting. The epitope recognized by the MAB was heat labile (60°, 30 min) and susceptible to enzymic digestion with pronase but was resistant to treatment with lipase, α -mannosidase, glucose oxidase, and endoglycosidase H. This heat-labile **peptide** epitope appears to be specific to **C. immitis**, as judged by the fact that the MAB was not reactive in immunoblots or ELISAs of histoplasmin or blastomycin.

L76 ANSWER 21 OF 25 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 18

ACCESSION NUMBER: 1989:455564 HCAPLUS

DOCUMENT NUMBER: 111:55564

TITLE: Induction of tumor necrosis factor alpha by spherules of **Coccidioides immitis**

AUTHOR(S): Slagle, David C.; Cox, Rebecca A.; Kuruganti, Uma

CORPORATE SOURCE: Dep. Res. Immunol., San Antonio State Chest Hosp., San Antonio, TX, 78223, USA

SOURCE: Infection and Immunity (1989), 57(7), 1916-21
CODEN: INFIBR; ISSN: 0019-9567

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The cytokine tumor necrosis factor α (TNF- α) functions as an immunomodulatory **protein** and as a mediator of cachexia. Here, the viable or formalin-killed spherules of **C. immitis** induced the secretion of TNF- α by peritoneal-exudate cells from BALB/c mice. The identification of the cytokine as TNF- α was based on its lytic activity against the TNF- α -sensitive Ls murine fibrosarcoma cell line but not the TNF- α -resistant LR cell line, its neutralization by rabbit anti-TNF- α , and its secretion by peritoneal cells having characteristics of macrophages. The induction of TNF- α was attributable to spherules and not to contaminating lipopolysaccharide (endotoxin), as evidenced by the finding that polymyxin B, a reagent that blocks the TNF- α -inducing component of lipopolysaccharide, did not negate the production of TNF- α in response to spherules, whereas pretreatment of spherules with hyperimmune goat antiserum to spherulin neutralized the induction of TNF- α by these cells. The demonstration that **C. immitis** activates macrophages to secrete TNF- α in vitro is a new finding and warrants studies to determine whether this cytokine is produced during active **coccidioidomycosis**.

L76 ANSWER 22 OF 25 MEDLINE on STN

DUPLICATE 19

Searcher : Shears 308-4994

10/081935

ACCESSION NUMBER: 88085464 MEDLINE
DOCUMENT NUMBER: 88085464 PubMed ID: 3335400
TITLE: Induction and expression of cell-mediated immune responses in inbred mice infected with **Coccidioides immitis**.
AUTHOR: Cox R A; Kennell W; Boncyk L; Murphy J W
CORPORATE SOURCE: Department of Research Immunology, San Antonio State Chest Hospital, Texas 78223.
CONTRACT NUMBER: AI23555 (NIAID)
SOURCE: INFECTION AND IMMUNITY, (1988 Jan) 56 (1) 13-7.
Journal code: 0246127. ISSN: 0019-9567.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 198802
ENTRY DATE: Entered STN: 19900305
Last Updated on STN: 19970203
Entered Medline: 19880209

AB Comparisons of the course of **coccidioidomycosis** in two strains of inbred mice established that BALB/c mice are significantly more susceptible to pulmonary infection with **Coccidioides immitis** than are DBA/2 mice. The susceptibility of BALB/c mice does not reside in their inability to mount a delayed-type hypersensitivity response to **C. immitis** antigen. That is, BALB/c mice manifested footpad hypersensitivity to **coccidioidin** early during the course of disease, to a level comparable to that of DBA/2 mice. In contrast to the more resistant DBA/2 mouse strain, however, BALB/c mice developed anergy by day 15 postinfection. Suppression of the delayed-type hypersensitivity response was not specific for **C. immitis** antigen, as evidenced by the finding that BALB/c mice immunized with mycobacterial purified **protein** derivative prior to infection with **C. immitis** were suppressed in their footpad response to mycobacterial antigen at day 15 postinfection. Taken together, these results establish that genetically determined susceptibility to this fungus is associated with an acquired suppression of cell-mediated immune reactivity.

L76 ANSWER 23 OF 25 MEDLINE on STN DUPLICATE 20
ACCESSION NUMBER: 87193089 MEDLINE
DOCUMENT NUMBER: 87193089 PubMed ID: 3552984
TITLE: Serum-mediated suppression of lymphocyte transformation responses in **coccidioidomycosis**.
AUTHOR: Cox R A; Pope R M
CONTRACT NUMBER: AI 18761 (NIAID)
SOURCE: INFECTION AND IMMUNITY, (1987 May) 55 (5) 1058-62.
Journal code: 0246127. ISSN: 0019-9567.
PUB. COUNTRY: United States
DOCUMENT TYPE: Journal; Article; (JOURNAL ARTICLE)
LANGUAGE: English
FILE SEGMENT: Priority Journals
ENTRY MONTH: 198706
ENTRY DATE: Entered STN: 19900303
Last Updated on STN: 19970203
Entered Medline: 19870605

10/081935

AB Lymphocyte transformation (LT) responses to **coccidioidin** (CDN) and spherulin were suppressed in 11 (73%) of 15 patients with active **coccidioidomycosis** when their mononuclear cells were assayed in autologous serum as compared to serum from healthy, CDN skin test-positive subjects. Suppressed LT responses were specific for **Coccidioides immitis** antigens in 7 (64%) of the 11 patients. Immunoaffinity chromatography of patient sera with *Staphylococcus* **protein A** adsorbed the suppressor component(s) and thereby established that suppression was attributed to immunoglobulin G, either alone or complexed with antigen. The possibility that suppression was mediated by immune complexes was examined by adding complexes formed in vivo or in vitro to mononuclear cell cultures of healthy CDN-reactive persons before LT assays. Although complexes prepared in this manner were reactive in an enzyme-linked immunosorbent assay designed to detect **Coccidioides** antigen-specific immune complexes, no suppression of LT responses was observed. We conclude that serum-mediated suppression of LT responses in **coccidioidomycosis** is attributed to monomeric and not immune-complexed immunoglobulin G antibody.

L76 ANSWER 24 OF 25 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN

ACCESSION NUMBER: 1984:33723 BIOSIS
DOCUMENT NUMBER: PREV198426033723; BR26:33723
TITLE: CHEMICAL COMPOSITION OF A MYCELIAL AND SPHERULE CELL WALL ANTIGEN FROM **COCCIDIODES-IMMITIS**.
AUTHOR(S): OLSBERG C A [Reprint author]; **COX R A**
CORPORATE SOURCE: UNIV TEX HEALTH SCI CENT, SAN ANTONIO, TX, USA
SOURCE: Abstracts of the Annual Meeting of the American Society for Microbiology, (1983) Vol. 83, pp. F6. Meeting Info.: 83RD ANNUAL MEETING OF THE AMERICAN SOCIETY FOR MICROBIOLOGY, NEW ORLEANS, LA., USA, MAR. 6-11, 1983. ABSTR ANNU MEET AM SOC MICROBIOL. CODEN: ASMACK. ISSN: 0094-8519.
DOCUMENT TYPE: Conference; (Meeting)
FILE SEGMENT: BR
LANGUAGE: ENGLISH

L76 ANSWER 25 OF 25 BIOSIS COPYRIGHT 2003 BIOLOGICAL ABSTRACTS INC. on STN

ACCESSION NUMBER: 1978:79432 BIOSIS
DOCUMENT NUMBER: PREV197815022932; BR15:22932
TITLE: CHEMICAL COMPOSITION OF AN ALKALI SOLUBLE WATER SOLUBLE ANTIGEN FROM **COCCIDIODES-IMMITIS**.
AUTHOR(S): FLEMING W H III; LAND G A; **COX R A**
SOURCE: Abstracts of the Annual Meeting of the American Society for Microbiology, (1978) Vol. 78.1978, pp. 314. CODEN: ASMACK. ISSN: 0094-8519.
DOCUMENT TYPE: Article
FILE SEGMENT: BR
LANGUAGE: Unavailable

FILE 'HOME' ENTERED AT 09:51:03 ON 04 DEC 2003

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 13, 2003, 12:34:43 ; Search time 214 Seconds
(without alignments)
274.896 Million cell updates/sec

Title: US-10-081-935-2

Perfect score: 82

Sequence: 1 MFQSHALIALVAAGLASA 18

Scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+.p2n.model -DEV=xlh

-Q=/cgn2_1/USPTO.spool/US10081935/runat_12112003_144408_3740/app.query.fasta_1.199

-DB=Published Applications NA -QFMT=fastap -SUPFI=rnpb -MINMATCH=0.1

-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=pct -THR_MAX=100

-THR_MIN=0 -ALIGN=0 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US10081935_@CGN_1_164_@runat_12112003_144408_3740

-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100

-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FCGAPOP=6 -FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:

3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:

13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:

14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:

15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:

16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:

17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	46	56.1	168	12	US-10-029-386-16850, A

C	2	46	56.1	569	12	US-10-029-386-3150	Sequence 3150, Ap
3	44	53.7	548	12	US-10-027-632-95893	Sequence 95893, A	
4	44	53.7	548	12	US-10-027-632-305915	Sequence 305915, A	
5	44	53.7	548	13	US-10-027-632-95893	Sequence 95893, A	
6	44	53.7	548	13	US-10-027-632-305915	Sequence 305915, A	
7	44	53.7	9372	12	US-10-293-582-2	Sequence 2, Appli	
8	43	52.4	549	11	US-09-991-936-889	Sequence 889, App	
9	43	52.4	549	11	US-09-991-936-889	Sequence 889, App	
10	43	52.4	720	10	US-09-894-844-66	Sequence 1638, Ap	
11	43	52.4	11115	12	US-09-769-734-49	Sequence 66, Appl	
12	42	51.2	2632	13	US-10-108-605-36	Sequence 49, Appl	
13	42	51.2	3042	10	US-09-712-363-68	Sequence 36, Appl	
14	41	50.0	251	14	US-10-198-846-12310	Sequence 12310, A	
15	41	50.0	498	11	US-09-991-936-975	Sequence 975, App	
16	41	50.0	609	10	US-09-738-628-3337	Sequence 3337, Ap	
17	41	50.0	894	14	US-10-198-846-4458	Sequence 4458, Ap	
18	41	50.0	1218	14	US-10-137-473-3	Sequence 3, Appli	
19	41	50.0	1722	14	US-10-156-761-4387	Sequence 4387, Ap	
20	41	50.0	2123	11	US-09-866-050A-584	Sequence 584, App	
21	41	50.0	2123	14	US-10-152-661-584	Sequence 584, App	
22	41	50.0	2484	14	US-10-156-761-918	Sequence 918, Appl	
23	41	50.0	3309400	10	US-09-738-626-1	Sequence 1, Appli	
24	41	50.0	9025608	14	US-10-156-761-1	Sequence 5360, Ap	
25	40	48.8	228	9	US-09-294-093B-5360	Sequence 70, Appl	
26	40	48.8	309	12	US-10-190-305A-71	Sequence 71, Appl	
27	40	48.8	309	12	US-10-190-305A-71	Sequence 66199, A	
28	40	48.8	406	12	US-10-027-632-66199	Sequence 66199, A	
29	40	48.8	406	13	US-10-027-632-66199	Sequence 4623, Ap	
30	40	48.8	436	10	US-09-867-701-4623	Sequence 36340, A	
31	40	48.8	445	11	US-09-918-995-36266	Sequence 36340, A	
32	40	48.8	467	12	US-10-027-632-36340	Sequence 36340, A	
33	40	48.8	467	13	US-10-027-632-36340	Sequence 36340, A	
34	40	48.8	509	12	US-10-027-632-42872	Sequence 42872, A	
35	40	48.8	509	13	US-10-027-632-42872	Sequence 42872, A	
36	40	48.8	512	12	US-10-027-632-281505	Sequence 281505, A	
37	40	48.8	512	13	US-10-027-632-281505	Sequence 281505, A	
38	40	48.8	513	11	US-09-918-995-30945	Sequence 30945, A	
39	40	48.8	622	12	US-10-027-632-185077	Sequence 185077, A	
40	48.8	622	12	US-10-027-632-185078	Sequence 185078, A		
41	40	48.8	622	13	US-10-027-632-185079	Sequence 185079, A	
42	40	48.8	622	13	US-10-027-632-185078	Sequence 185078, A	
43	40	48.8	622	13	US-10-027-632-185079	Sequence 185079, A	
44	40	48.8	622	13	US-10-027-632-185079	Sequence 203917, A	
45	40	48.8	637	12	US-10-027-632-203917	Sequence 203917, A	

ALIGNMENTS

RESULT 1

US-10-029-386-16850/c

; Sequence 16850, Application US/10029386

; Publication No. US20030194704A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO

; FILE REFERENCE: AEMICA-X-2

; CURRENT APPLICATION NUMBER: US/10/029,386

; NUMBER OF SEQ ID NOS: 34288

; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 16850

; LENGTH: 168

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO CHR1.1

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9

; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1

```
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: SWISSPROT HIT: Q93008, EVALUE 5.00e-20
; OTHER INFORMATION: NT HIT: gi14764091, EVALUE 9.00e-71
; OTHER INFORMATION: EST_HUMAN HIT: AW996853.1, EVALUE 5.00e-73
US-10-029-386-16850
Alignment Scores:
Pred. No.: 2.37 Length: 168
Score: 46.00 Matches: 9
Percent Similarity: 76.47% Conservative: 4
Best Local Similarity: 52.94% Mismatches: 4
Query Match: 56.10% Indels: 0
DB: 12 Gaps: 0

US-10-081-935-2 (1-18) x US-10-029-386-16850 (1-168)
Qy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSer 17
Db 110 CTTCAACACAATCATGCCCTAGTTACTTTGGTAGCAGAAAACCTTGCACACT 60
RESULT 2
US-10-029-386-3150/c
; Sequence 3150, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Shatton G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: AROMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3150
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR.X.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: SWISSPROT HIT: Q93008, EVALUE 2.00e-19
; OTHER INFORMATION: EST_HUMAN HIT: AW996853.1, EVALUE 5.00e-79
; OTHER INFORMATION: NT HIT: gi14764084, EVALUE 3.00e-70
US-10-029-386-3150
Alignment Scores:
Pred. No.: 10.8 Length: 569
Score: 46.00 Matches: 9
Percent Similarity: 76.47% Conservative: 4
Best Local Similarity: 52.94% Mismatches: 4
Query Match: 56.10% Indels: 0
DB: 12 Gaps: 0

US-10-081-935-2 (1-18) x US-10-029-386-3150 (1-569)
Qy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSer 17
Db 276 CTTCAACACAATCATGCCCTAGTTACTTTGGTAGCAGAAAACCTTGCACACT 226
RESULT 3
US-10-027-632-95893
; Sequence 95893, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```

```
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 95893
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-95893
Alignment Scores:
Pred. No.: 25.6 Length: 548
Score: 44.00 Matches: 8
Percent Similarity: 83.33% Conservative: 2
Best Local Similarity: 66.67% Mismatches: 2
Query Match: 53.66% Indels: 0
DB: 12 Gaps: 0

US-10-081-935-2 (1-18) x US-10-027-632-95893 (1-548)
Qy 3 PheSerHisAlaLeuIleAlaLeuValAlaAlaGly 14
Db 320 TTTTCCCATATGATGATAGCATTCACTAATACTGCTGGG 355
RESULT 4
US-10-027-632-305915
; Sequence 305915, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 305915
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-305915
```

Alignment Scores:
Pred. No.: 25.6 Length: 548
Score: 44.00 Matches: 8
Percent Similarity: 83.33% Conservative: 2
Best Local Similarity: 66.67% Mismatches: 2
Query Match: 53.66% Indels: 0
DB: 12 Gaps: 0

US-10-081-935-2 (1-18) x US-10-027-632-305915 (1-548)

QY 3 PheSerHisAlaLeuIleAlaLeuValAlaAlaGly 14
||||| :|||:|||||
DB 320 TTTTCCCATATGATAGCACTCATAACTGCTGGG 355

RESULT 5

US-10-027-632-95893
; Sequence 95893, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 95893
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-95893

Alignment Scores:
Pred. No.: 25.6 Length: 548
Score: 44.00 Matches: 8
Percent Similarity: 83.33% Conservative: 2
Best Local Similarity: 66.67% Mismatches: 2
Query Match: 53.66% Indels: 0
DB: 12 Gaps: 0

US-10-081-935-2 (1-18) x US-10-027-632-95893 (1-548)

QY 3 PheSerHisAlaLeuIleAlaLeuValAlaAlaGly 14
||||| :|||:|||||
DB 320 TTTTCCCATATGATAGCACTCATAACTGCTGGG 355

RESULT 6

US-10-027-632-305915
; Sequence 305915, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 305915
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-305915

Alignment Scores:
Pred. No.: 25.6 Length: 548
Score: 44.00 Matches: 8
Percent Similarity: 83.33% Conservative: 2
Best Local Similarity: 66.67% Mismatches: 2
Query Match: 53.66% Indels: 0
DB: 12 Gaps: 0

US-10-081-935-2 (1-18) x US-10-027-632-305915 (1-548)

QY 3 PheSerHisAlaLeuIleAlaLeuValAlaAlaGly 14
||||| :|||:|||||
DB 320 TTTTCCCATATGATAGCACTCATAACTGCTGGG 355

RESULT 7

US-10-293-582-2
; Sequence 2, Application US/10293582
; Publication No. US20030175253A1
; GENERAL INFORMATION:
; APPLICANT: Akil, Huda
; APPLICANT: Bunney, William E.
; APPLICANT: Burke, Sharon
; APPLICANT: Choudary, Prabhakara V.
; APPLICANT: Cox, David R.
; APPLICANT: Evans, Simon
; APPLICANT: Jones, Edward G.
; APPLICANT: Li, Jun
; APPLICANT: Lopez, Juan F.
; APPLICANT: The Trustees of The Leland Stanford Junior University
; TITLE OF INVENTION: Compositions and Methods for Diagnosing and Treating
; TITLE OF INVENTION: Mental Disorders
; FILE REFERENCE: 020885-0002100S
; CURRENT APPLICATION NUMBER: US/10/293,582
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/339,252
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 9372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: chromosome Y ubiquitin-specific cysteine protease
US-10-293-582-2

Alignment Scores:
Pred. No.: 881 Length: 9372
Score: 44.00 Matches: 9
Percent Similarity: 76.47% Conservative: 4
Best Local Similarity: 52.94% Mismatches: 4
Query Match: 53.66% Indels: 0


```
DB: 12 Gaps: 0
US-10-081-935-2 (1-18) x US-10-293-582-2 (1-9372)
Qy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSer 17
Db 2764 CTTCAACAAATCATGCTTTAGTTACTTTGGTAGCAGAAACCTTGCACAC 2814

RESULT 8
US-09-991-936-889
; Sequence 889, Application US/09991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
; FILE REFERENCE: FC-6-C1
; CURRENT APPLICATION NUMBER: US/09/991,936
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US/09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1959
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 889
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-991-936-889

Alignment Scores:
Pred. No.: 40.4 Length: 549
Score: 43.00 Matches: 9
Percent Similarity: 72.22% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 52.44% Indels: 0
DB: 11 Gaps: 0

US-10-081-935-2 (1-18) x US-09-991-936-889 (1-549)
Qy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 21 ATGAAATTCGACGTAGCAATTTGGCCCTGCTGTGTTGGCATCAGCT 74

RESULT 9
US-09-991-936-1638
; Sequence 1638, Application US/09991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
; FILE REFERENCE: FC-6-C1
; CURRENT APPLICATION NUMBER: US/09/991,936
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US/09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1959
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1638
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-991-936-1638
```

```
Alignment Scores:
Pred. No.: 40.4 Length: 549
Score: 43.00 Matches: 9
Percent Similarity: 72.22% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 52.44% Indels: 0
DB: 11 Gaps: 0

US-10-081-935-2 (1-18) x US-09-991-936-1638 (1-549)
Qy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 21 ATGAAATTCGACGTAGCAATTTGGCCCTGCTGTGTTGGCATCAGCT 74

RESULT 10
US-09-894-844-66
; Sequence 66, Application US/09894844
; Patent No. US20020176873A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of
; FILE REFERENCE: STAN102CON
; CURRENT APPLICATION NUMBER: US/09/894,844
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/318,191
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/097,936
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 720
; TYPE: DNA
; ORGANISM: M. tuberculosis
US-09-894-844-66

Alignment Scores:
Pred. No.: 56.6 Length: 720
Score: 43.00 Matches: 10
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 90.91% Mismatches: 0
Query Match: 52.44% Indels: 0
DB: 10 Gaps: 0

US-10-081-935-2 (1-18) x US-09-894-844-66 (1-720)
Qy 8 IleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 25 ATCGCCCTGGTGGCCGCCGCTTGGCCGCCGA 57

RESULT 11
US-09-769-734-49/C
; Sequence 49, Application US/09769734
; Publication No. US20030143666A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia BioSciences Inc.
; TITLE OF INVENTION: Genetic Locus for Everninomicin Biosynthesis
; FILE REFERENCE: PA 005-US
; CURRENT APPLICATION NUMBER: US/09/769,734
; CURRENT FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 49
; LENGTH: 11115
; TYPE: DNA
; ORGANISM: M. carbonacea
; FEATURE:
; NAME/KEY: misc_feature
```

LOCATION: (8)..(1207)
; OTHER INFORMATION: ORF 41 (positive strandedness)
; NAME/KEY: misc_feature
LOCATION: (1213)..(2331)
; OTHER INFORMATION: ORF 42 (positive strandedness)
; NAME/KEY: misc_feature
LOCATION: (2364)..(3611)
; OTHER INFORMATION: ORF 43 (positive strandedness)
; NAME/KEY: misc_feature
LOCATION: (3623)..(4243)
; OTHER INFORMATION: ORF 44 (positive strandedness)
; NAME/KEY: misc_feature
LOCATION: (4149)..(5177)
; OTHER INFORMATION: ORF 45 (positive strandedness)
; NAME/KEY: misc_feature
LOCATION: (5177)..(6094)
; OTHER INFORMATION: ORF 46 (negative strandedness)
; NAME/KEY: misc_feature
LOCATION: (6271)..(7824)
; OTHER INFORMATION: ORF 47 (negative strandedness)
; NAME/KEY: misc_feature
LOCATION: (7903)..(8760)
; OTHER INFORMATION: ORF 48 (negative strandedness)
; NAME/KEY: misc_feature
LOCATION: (8781)..(9800)
; OTHER INFORMATION: ORF 49 (negative strandedness)
US-09-769-734-49

Alignment Scores:
Pred. No.: 1.71e+03 Length: 11115
Score: 43.00 Matches: 12
Percent Similarity: 63.64% Conservative: 2
Best Local Similarity: 54.55% Mismatches: 2
Query Match: 52.44% Indels: 6
DB: 12 Gaps: 1

US-10-081-935-2 (1-18) x US-09-769-734-49 (1-11115)

Qy 1 MetGlnPheSerHisAlaLeuIleAla-----LeuValAlaAlaGly 14
Db 7824 ATGAGACATGTCATGCCCTGTGGCGCTGGACACCGGACCGCTGTGGCGCGGC 7765
Qy 15 LeuAla 16
Db 7764 CTGGCC 7759

RESULT 12
US-10-108-605-36
; Sequence 36, Application US/10108605
; Publication No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kandari, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 2632
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-108-605-36

Alignment Scores:
Pred. No.: 448 Length: 2632
Score: 42.00 Matches: 8
Percent Similarity: 91.67% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 1
Query Match: 51.22% Indels: 0
DB: 13 Gaps: 0

US-10-081-935-2 (1-18) x US-10-108-605-36 (1-2632)

Qy 4 SerHisAlaLeuIleAlaLeuValAlaAlaGlyLeu 15
Db 1870 TCCCACTCCGCTGTTCCCTCGTTCGCCGAGGACTT 1905

RESULT 13

US-09-712-363-68
; Sequence 68, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-03201
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 3042
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-68

Alignment Scores:
Pred. No.: 536 Length: 3042
Score: 42.00 Matches: 9
Percent Similarity: 82.35% Conservative: 5
Best Local Similarity: 52.94% Mismatches: 3
Query Match: 51.22% Indels: 0
DB: 10 Gaps: 0

US-10-081-935-2 (1-18) x US-09-712-363-68 (1-3042)

Qy 2 GlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 174 CGTTTGTCTGTTGCCGTGCTGCGATCGGCGCGCTGCGGAGCT 224

RESULT 14

US-10-198-846-12310/c
; Sequence 12310, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James

```
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12310
; LENGTH: 251
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-12310

Alignment Scores:
Pred. No.: 37.6 Length: 251
Score: 41.00 Matches: 9
Percent Similarity: 71.43% Conservative: 1
Best Local Similarity: 64.23% Mismatches: 4
Query Match: 50.00% Indels: 0
DB: 14 Gaps: 0

US-10-081-935-2 (1-18) x US-10-198-846-12310 (1-251)

Qy 5 HisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 122 CACGCGATGATCGTCTTTCGCCGCCGCGCTGCTCGGCT 81

RESULT 15
US-09-991-936-975/c
; Sequence 975, Application US/09991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-6-C1
; CURRENT APPLICATION NUMBER: US/09/991,936
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US/09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1959
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 975
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-991-936-975

Alignment Scores:
Pred. No.: 88.4 Length: 498
Score: 41.00 Matches: 11
Percent Similarity: 68.75% Conservative: 0
Best Local Similarity: 68.75% Mismatches: 3
Query Match: 50.00% Indels: 2
DB: 11 Gaps: 1

US-10-081-935-2 (1-18) x US-09-991-936-975 (1-498)
```

```
Qy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAla 16
Db 445 ATGGATTTCCTCAC-----ATAACGTTAGTTGCAAAAGGATTAGCT 404

RESULT 16
US-09-738-626-3337
; Sequence 3337, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3337
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3337

Alignment Scores:
Pred. No.: 114 Length: 609
Score: 41.00 Matches: 8
Percent Similarity: 76.47% Conservative: 5
Best Local Similarity: 47.06% Mismatches: 4
Query Match: 50.00% Indels: 0
DB: 10 Gaps: 0

US-10-081-935-2 (1-18) x US-09-738-626-3337 (1-609)

Qy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSer 17
Db 400 CTTACATCGGATACGCCATCATCGTATCCGCCGACGAGGTCTCGCGCT 450

RESULT 17
US-10-198-846-4458/c
; Sequence 4458, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4458
; LENGTH: 894
; TYPE: DNA
```

Df 480 CTGCGCTTTGGACCAAGCCTCGTGTCTTCCTTGGAATGGGTCTGGTCACAGCA 53

RESULT 19
US-10-156-761-4387/c
; Sequence 4387, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAI, YOSHIIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4387
; LENGTH: 1722
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1722)
US-10-156-761-4387

Alignment Scores:
Pred. No.: 415 Length: 1722
Score: 41.00 Matches: 9
Percent Similarity: 86.67% Conservative: 4
Best Local Similarity: 60.00% Mismatches: 2
Query Match: 50.00% Indels: 0
DB: 14 Gaps: 0

US-10-081-935-2 (1-18) x US-10-156-761-4387 (1-1722)

Oy 1 MetGlnPheSerHisAlaLeuIfleAlaLeuValAlaLaAglyLeu 15
 :::|::|::|::|::|::|::|::|::|::|::|::|::|
Df 492 GTCCAGGACGCAGTGTCCACTGTCTGCCCGGGCGCTT 448

RESULT 20
US-09-866-050A-584
; Sequence 584, Application US/09866050A
; Publication No. US2003004071A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishnanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE REFERENCE: 11000.111c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 584
; LENGTH: 2123
; TYPE: DNA
; ORGANISM: Mouse
US-09-866-050A-584

Alignment Scores:
Pred. No.: 539 Length: 2123
Score: 41.00 Matches:

Percent Similarity: 66.67% Conservatives: 4
Best Local Similarity: 44.44% Mismatches: 6
Query Match: 50.00% Indels: 0
DB: 11 Gaps: 0

US-10-081-935-2 (1-18) x US-09-866-050A-584 (1-2123)

Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 870 CTGCTTTGAGCCACGCTGCTGCTTCTCTGGCAATGGTCTGTCACAGCA 923

RESULT 21

US-10-152-661-584
; Sequence 584, Application US/10152661
; Publication No. US20030022835A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; FILE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c5
; CURRENT APPLICATION NUMBER: US/10/152,661
; CURRENT FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: 09/866,050
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/221,232
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/206,650
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 09/312,283
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/N299/00051
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 09/188,930
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: 09/069,726
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 584
; LENGTH: 2123
; TYPE: DNA
; ORGANISM: Mouse

Alignment Scores:
Pred. No.: 539 Length: 2123
Score: 41.00 Matches: 8
Percent Similarity: 66.67% Conservatives: 4
Best Local Similarity: 44.44% Mismatches: 6
Query Match: 50.00% Indels: 0
DB: 14 Gaps: 0

US-10-081-935-2 (1-18) x US-10-152-661-584 (1-2123)

Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 870 CTGCTTTGAGCCACGCTGCTGCTTCTCTGGCAATGGTCTGTCACAGCA 923

RESULT 22

US-10-156-761-918
; Sequence 918, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 918
; LENGTH: 2484
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2484)
US-10-156-761-918

Alignment Scores:
Pred. No.: 655 Length: 2484
Score: 41.00 Matches: 8
Percent Similarity: 70.59% Conservatives: 4
Best Local Similarity: 47.06% Mismatches: 5
Query Match: 50.00% Indels: 0
DB: 14 Gaps: 0

US-10-081-935-2 (1-18) x US-10-156-761-918 (1-2484)

Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSer 17
Db 1168 CTGCAACTGACACACCGCTTGTCTGCTGAGCGGGCTGCGCCGCC 1218

RESULT 23

US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1
Alignment Scores:
Pred. No.: 3e+06 Length: 3309400
Score: 41.00 Matches: 8
Percent Similarity: 76.47% Conservatives: 5
Best Local Similarity: 47.06% Mismatches: 4
Query Match: 50.00% Indels: 0

```
DB: 10 Gaps: 0
US-10-081-935-2 (1-18) x US-09-738-626-1 (1-3309400)
Qy 1 MetGlnPheSerHisAlaLeuValAlaGlyLeuAlaSer 17
Db 3218392 CTTGACATCGGATACGCCATCATCGGTATCGCGCAGCAGGTCTCGCGCT 3218442
RESULT 24
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, HARUO
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 4.25e+06 Length: 9025608
Score: 41.00 Matches: 8
Percent Similarity: 70.59% Conservative: 4
Best Local Similarity: 47.06% Mismatches: 5
Query Match: 50.00% Indels: 0
DB: 14 Gaps: 0
US-10-081-935-2 (1-18) x US-10-156-761-1 (1-9025608)
Qy 1 MetGlnPheSerHisAlaLeuValAlaGlyLeuAlaSer 17
Db 1126427 CTGCAACTGACACACCGCTTGCTGCACCTTCGTCGAGCGGGGCTGGCGCC 1126477
RESULT 25
US-09-294-093B-5360/c
; Sequence 5360, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Laigudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 5360
; LENGTH: 228
; TYPE: DNA
; ORGANISM: Zea mays
```

```
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700356446H1
US-09-294-093B-5360
Alignment Scores:
Pred. No.: 52.5 Length: 228
Score: 40.00 Matches: 8
Percent Similarity: 84.62% Conservative: 3
Best Local Similarity: 61.54% Mismatches: 2
Query Match: 48.78% Indels: 0
DB: 9 Gaps: 0
US-10-081-935-2 (1-18) x US-09-294-093B-5360 (1-228)
Qy 1 MetGlnPheSerHisAlaLeuValAlaGlyLeuAlaSer 13
Db 153 ATCCAATTGAATCATGCCCTTCTGCTCTTGTCAAAGCT 115
RESULT 26
US-10-190-305A-70/c
; Sequence 70, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: TatC22Exon1_2_Tvl_C_ZAopt
US-10-190-305A-70
Alignment Scores:
Pred. No.: 76.7 Length: 309
Score: 40.00 Matches: 8
Percent Similarity: 78.57% Conservative: 3
Best Local Similarity: 57.14% Mismatches: 3
Query Match: 48.78% Indels: 0
DB: 12 Gaps: 0
US-10-081-935-2 (1-18) x US-10-190-305A-70 (1-309)
Qy 5 HisAlaLeuValAlaGlyLeuAlaSerAla 18
Db 273 CACCTTCTTCTGCTCTCTCGTGGTGGTCTGCTCGCCGCG 232
RESULT 27
US-10-190-305A-71/c
; Sequence 71, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
```

```

US-10-027-632-66199/c
; Sequence 66199, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/219,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66199
; LENGTH: 406
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(406)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-66199

Alignment Scores:
Pred. No.: 108 Length: 406
Score: 40.00 Matches: 9
Percent Similarity: 75.00% Conservative: 3
Best Local Similarity: 56.25% Mismatches: 4
Query Match: 48.78% Indels: 0
DB: 13 Gaps: 0

US-10-081-935-2 (1-18) x US-10-027-632-66199 (1-406)

QY 3 PheSerHisAlaLeuLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 211 TTTTCGAATGCTCTTATGCTTTACTTGGGGTGGTCATTGCTCAGCT 164

RESULT 30
US-09-867-701-4623
; Sequence 4623, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4623
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-4623

Alignment Scores:
Pred. No.: 118 Length: 436
Score: 40.00 Matches: 8

```

```

Percent Similarity: 71.43% Conservative: 2
Best Local Similarity: 57.14% Mismatches: 4
Query Match: 48.78% Indels: 0
DB: 10 Gaps: 0

US-10-081-935-2 (1-18) x US-09-867-701-4623 (1-436)

QY 1 MetGlnPheSerHisAlaLeuValAlaLeuValAlaAlaGly 14
   ||| :::||||| ::| ||||| |||||
Db 371 ATGACATACAGTCATCGCACCTGGAGCTAGTCTGGCTGGG 412

RESULT 31
US-09-918-995-36266/c
; Sequence 36266, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: HySeq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36266
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(445)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-36266

Alignment Scores:
Pred. No.: 121 Length: 445
Score: 76.92% Matches: 9
Percent Similarity: 76.92% Conservative: 1
Best Local Similarity: 69.23% Mismatches: 3
Query Match: 48.78% Indels: 0
DB: 11 Gaps: 0

US-10-081-935-2 (1-18) x US-09-918-995-36266 (1-445)

QY 4 SerHisAlaLeuValAlaLeuValAlaAlaGlyLeuAla 16
   ||||| ||||| ||||| ||||| |||||
Db 390 AGCCATCTCTGATCTCTTGAAGCGGACGAGCATCGCT 352

RESULT 32
US-10-027-632-36340/c
; Sequence 36340, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single N
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28

```


Oy 3 PheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18
|||||:|||||:|||||:|||||:|||||
Db 211 TTTTGAATGCTCTTATGGCTTTACTTGGGGTGGTGCATGCTCAGCT 164

RESULT 34

US-10-027-632-42872
; Sequence 42872, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 42872
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-42872

Alignment Scores:
Pred. No.: 143 Length: 509
Score: 40.00 Matches: 7
Percent Similarity: 90.91% Conservative: 3
Best Local Similarity: 63.64% Mismatches: 1
Query Match: 48.78% Indels: 0
DB: 12 Gaps: 0

US-10-081-935-2 (1-18) x US-10-027-632-42872 (1-509)

Oy 3 PheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 13
|||||:|||||:|||||:|||||:|||||
Db 369 TTTTCTCATTCCTTAGTTCGAATAGTAGGGGCC 401

RESULT 35

US-10-027-632-42872
; Sequence 42872, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 42872
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-42872

Alignment Scores:
Pred. No.: 143 Length: 509
Score: 40.00 Matches: 7
Percent Similarity: 90.91% Conservative: 3
Best Local Similarity: 63.64% Mismatches: 1
Query Match: 48.78% Indels: 0
DB: 12 Gaps: 0

US-10-081-935-2 (1-18) x US-10-027-632-42872 (1-509)

Oy 3 PheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 13
|||||:|||||:|||||:|||||:|||||
Db 369 TTTTCTCATTCCTTAGTTCGAATAGTAGGGGCC 401

RESULT 36

US-10-027-632-281505
; Sequence 281505, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 281505
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-281505

Alignment Scores:
Pred. No.: 144 Length: 512
Score: 40.00 Matches: 7
Percent Similarity: 80.00% Conservative: 5
Best Local Similarity: 46.67% Mismatches: 3
Query Match: 48.78% Indels: 0
DB: 12 Gaps: 0

US-10-081-935-2 (1-18) x US-10-027-632-281505 (1-512)

Oy 4 SerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18
|||||:|||||:|||||:|||||:|||||
Db 112 TCCCATGCACTAGTGGCCATGATCATTTATTTGGCAGCTCTCCAGT 156

RESULT 37

```
US-10-027-632-281505
; Sequence 281505, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 281505
; LENGTH: 512
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-281505
Alignment Scores:
Pred. No.: 144      Length: 512
Score: 40.00      Matches: 7
Percent Similarity: 80.00%      Conservative: 5
Best Local Similarity: 46.67%      Mismatches: 3
Query Match: 48.78%      Indels: 0
DB: 13      Gaps: 0

US-10-081-935-2 (1-18) x US-10-027-632-281505 (1-512)
Qy 4 SerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSer 18
Db 112 TCCCATGCACCTAGTGGCCATGATCATTTATTGGCAGCTCCTCCACT 156

US-09-918-995-30945
; Sequence 30945, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30945
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(513)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-30945
Alignment Scores:
Pred. No.: 144      Length: 513
Score: 40.00      Matches: 8

US-10-027-632-185077
; Sequence 185077, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 185077
; LENGTH: 622
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-185077
Alignment Scores:
Pred. No.: 183      Length: 622
Score: 40.00      Matches: 8
Percent Similarity: 83.33%      Conservative: 2
Best Local Similarity: 66.67%      Mismatches: 2
Query Match: 48.78%      Indels: 0
DB: 12      Gaps: 0

US-10-081-935-2 (1-18) x US-10-027-632-185077 (1-622)
Qy 3 PheSerHisAlaLeuIleAlaLeuValAlaAlaGly 14
Db 436 TTCTCCCACTCCTAGTGGCCATGATTTGGAGGCCAATGTT 471

US-10-027-632-185078
; Sequence 185078, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
```

```

; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 185078
; LENGTH: 622
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-185078

```

```

Alignment Scores:
Pred. No.: 183 Length: 622
Score: 40.00 Matches: 8
Percent Similarity: 83.33% Conservative: 2
Best Local Similarity: 66.67% Mismatches: 2
Query Match: 48.78% Indels: 0
DB: 12 Gaps: 0

```

US-10-081-935-2 (1-18) x US-10-027-632-185078 (1-622)

```

Qy      3 PheSerHisAlaLeuIleAlaLeuValAlaAlaGly 14
         |||||:::|||||:::|||||
Db      436 TTCTCCCACTCACTGGTAGCATTGGAGGCCAATGGT 471

```

Search completed: November 13, 2003, 14:12:16
Job time : 984 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 21:23:48 ; Search time 242 Seconds
(without alignments)
580.045 Million cell updates/sec

Title: US-10-081-935-1
Perfect score: 52
Sequence: 1 atgcagttcttcacgctct.....cgctgcggcctcgccagtg 52

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N:Geneseq, 19Jun03:.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	25.6	49.2	1006	23 AAS59682 Propionibacterium
2	25.6	49.2	34088	23 AAS55566 Propionibacterium
3	25	48.1	336	22 AAD20424 Pseudomonas syring
C 4	25	48.1	1920	23 AAS54321 Pseudomonas aerugi
C 5	24.8	47.7	2016	23 ABL07659 Drosophila melanog
C 6	24.8	47.7	4095	23 ABL07658 Drosophila melanog
C 7	24.6	47.3	1566	22 AAS57171 DNA encoding Dros
C 8	24.6	47.3	1566	22 AAH49426 D. melanogaster pe

C 9	24.6	47.3	1566	23 ABL22999 Drosophila melanog
C 10	24.6	47.3	3772	22 AAS57170 DNA encoding Dros
C 11	24.6	47.3	3772	23 ABL22998 Drosophila melanog
C 12	24.4	46.9	695	21 AAF12891 Aspergillus oryzae
C 13	24.4	46.9	5484	23 AAS73006 DNA encoding novel
C 14	24.4	46.9	25913	22 AAK70862 Human immune/haema
C 15	24.4	46.9	198161	24 ABK83564 Human cDNA differe
C 16	24	46.2	2162	21 AAA94724 Rice auxin transpo
C 17	23.8	45.8	609	22 AAH68302 C glutamicum codin
C 18	23.8	45.8	732	22 AAF71300 Corynebacterium gl
C 19	23.8	45.8	7077	25 ABZ59094 E. maxima 230 kDa
C 20	23.8	45.8	7077	25 ABZ22977 Eimeria maxima 250
C 21	23.8	45.8	7987	25 ABZ22976 Eimeria maxima 250
C 22	23.8	45.8	7990	25 ABZ22984 E. maxima immunodo
C 23	23.8	45.8	49377	19 AAV05287 The soraphen biosy
C 24	23.8	45.8	58857	21 AAH58471 Nucleotide sequenc
C 25	23.8	45.8	309400	22 AAH68534 C glutamicum codin
C 26	23.6	45.4	1415	24 ABZ11723 Human polynucleoti
C 27	23.6	45.4	135638	25 ABX34289 S. atroolivaceus l
C 28	23.4	45.0	145	24 AAD38151 pCON3 deletion mut
C 29	23.4	45.0	450	24 AEN20845 Human ORFX polynuc
C 30	23.4	45.0	855	21 AA94737 Corn auxin transpo
C 31	23.4	45.0	2466	21 AAC55847 Mitomycin biosynth
C 32	23.4	45.0	2840	23 ABL13923 Drosophila melanog
C 33	23.4	45.0	5736	23 ABL08443 Drosophila melanog
C 34	23.4	45.0	6782	23 ABL13922 Drosophila melanog
C 35	23.4	45.0	9577	23 ABL08442 Drosophila melanog
C 36	23.4	45.0	18034	21 AAC55841 Complete Mitomycin
C 37	23.2	44.6	271	25 ABX30608 Human GDP-mannose
C 38	23.2	44.6	450	24 ABQ98305 Human ORF112 codin
C 39	23.2	44.6	867	23 ABL26959 Drosophila melanog
C 40	23.2	44.6	1563	24 ABL60268 Trametes versicolor
C 41	23.2	44.6	1899	16 AAT05530 Metalloprotease co
C 42	23.2	44.6	2052	16 AAT05529 p45 metalloproteas
C 43	23.2	44.6	2052	17 AAT40133 Fusarium oxysporum
C 44	23.2	44.6	2052	19 AAV28868 Fusarium oxysporum
C 45	23.2	44.6	5437	12 AAQ13369 Fuesaric acid resis

ALIGNMENTS

RESULT 1
AAS59682
ID AAS59682 standard; DNA; 1006 BP.

XX AAS59682;

AC AAS59682;

DT 13-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein encoding DNA #177.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant; ds.

XX Propionibacterium acnes.

OS WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

PA Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

```

XX  .
XX  WPI; 2001-616774/71.
XX  Propionibacterium acnes polypeptides and nucleic acids useful for
XX  PT vaccinating against and diagnosing infections, especially useful for
XX  PT treating acne vulgaris -
XX  XX
XX  Claim 1; SEQ ID No 177; 1069pp; English.
XX  XX
XX  Sequences AAS59506-AA59804 represent DNA molecules encoding
XX  CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
XX  CC associated DNA sequences are used in the treatment, prevention and
XX  CC diagnosis of medical conditions caused by P. acnes. The disorders include
XX  CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
XX  CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
XX  CC in infections of bone, joints and the central nervous system, however it
XX  CC is particularly involved in the inflammatory lesions associated with acne
XX  CC vulgaris. A method for detecting the presence or absence of P. acnes in a
XX  CC patient comprises contacting a sample with a binding agent that binds to
XX  CC the proteins of the invention and determining the amount of bound protein
XX  CC in the sample. The polypeptides may be used as antigens in the production
XX  CC of antibodies specific for P. acnes proteins. These antibodies can be
XX  CC used to downregulate expression and activity of P. acnes polypeptides and
XX  CC therefore treat P. acnes infections. The antibodies may also be used as
XX  CC diagnostic agents for determining P. acnes presence, for example, by
XX  CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
XX  CC polypeptides shown in AAU5636-AAU65646 and AAU67799-AAU67800.
XX  CC Note: The sequence data for this patent did not form part of the printed
XX  CC specification, but was obtained in electronic format directly from WIPO
XX  CC at ftp.wipo.int/pub/published_pct_sequences.
XX  XX
XX  SQ Sequence 1006 BP; 201 A; 352 C; 250 G; 199 T; 4 other;

Query Match      49.2%; Score 25.6; DB 23; Length 1006;
Best Local Similarity 70.8%; Pred. No. 26;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      3 GCAGTTCTCTCACGCTCTCATCGCTCTCGTGGCGGCTCGCCAG 50
Db      915 GCAGTTCACTCAATCCCTCGTCGATCTCTTTGAGACCAGCGTCGCGAG 962

RESULT 2
AAS59566
ID      AAS59566 standard; DNA; 34088 BP.
XX
XX  AAS59566;
AC
XX
XX  13-FEB-2002 (first entry)
DT
XX
XX  Propionibacterium acnes immunogenic protein encoding DNA #61.
DE
XX
XX  SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX  KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX  KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX  KW dermatological; osteopathic; neuroprotectant; ds.
XX
XX  Propionibacterium acnes.
OS
XX
XX  WO200181581-A2.
PN
XX
XX  01-NOV-2001.
PD
XX
XX  20-APR-2001; 2001WO-US12865.
PF
XX
XX  21-APR-2000; 2000US-199047P.
PR
XX  02-JUN-2000; 2000US-208841P.
PR
XX  07-JUL-2000; 2000US-216747P.
XX
XX  (CORI-) CORIXA CORP.
PA
XX
XX  Steiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

```

```

XX  WPI; 2001-616774/71.
XX  Propionibacterium acnes polypeptides and nucleic acids useful for
XX  PT vaccinating against and diagnosing infections, especially useful for
XX  PT treating acne vulgaris -
XX  XX
XX  Claim 1; SEQ ID No 61; 1069pp; English.
XX  XX
XX  Sequences AAS59506-AA59804 represent DNA molecules encoding
XX  CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
XX  CC associated DNA sequences are used in the treatment, prevention and
XX  CC diagnosis of medical conditions caused by P. acnes. The disorders include
XX  CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
XX  CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
XX  CC in infections of bone, joints and the central nervous system, however it
XX  CC is particularly involved in the inflammatory lesions associated with acne
XX  CC vulgaris. A method for detecting the presence or absence of P. acnes in a
XX  CC patient comprises contacting a sample with a binding agent that binds to
XX  CC the proteins of the invention and determining the amount of bound protein
XX  CC in the sample. The polypeptides may be used as antigens in the production
XX  CC of antibodies specific for P. acnes proteins. These antibodies can be
XX  CC used to downregulate expression and activity of P. acnes polypeptides and
XX  CC therefore treat P. acnes infections. The antibodies may also be used as
XX  CC diagnostic agents for determining P. acnes presence, for example, by
XX  CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
XX  CC polypeptides shown in AAU54556-AAU54928 and AAU67560-AAU67562.
XX  CC Note: The sequence data for this patent did not form part of the printed
XX  CC specification, but was obtained in electronic format directly from WIPO
XX  CC at ftp.wipo.int/pub/published_pct_sequences.
XX  XX
XX  SQ Sequence 34088 BP; 6991 A; 11002 C; 9453 G; 6639 T; 3 other;

Query Match      49.2%; Score 25.6; DB 23; Length 34088;
Best Local Similarity 70.8%; Pred. No. 40;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      3 GCAGTTCTCTCACGCTCTCATCGCTCTCGTGGCGGCTCGCCAG 50
Db      29931 GCAGTTCACTCAATCCCTCGTCGATCTCTTTGAGACCAGCGTCGCGAG 29978

RESULT 3
AAD20424
ID      AAD20424 standard; DNA; 336 BP.
XX
XX  AAD20424;
AC
XX
XX  03-JAN-2002 (first entry)
DT
XX
XX  Pseudomonas syringae pv. syringae (Psy) shcA DNA.
DE
XX
XX  Conserved Effector Loci; CEL; cytostatic; antibacterial; gene therapy;
XX  KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
XX  KW eukaryotic cell death; cancer; ds.
XX
XX  Pseudomonas syringae.
OS
XX
XX  Key      Location/Qualifiers
XX  CDS      1..336
XX  FT       /*tag= a
XX  FT       /product= "Psy shcA protein"
XX
XX  WO200175066-A2.
PN
XX
XX  11-OCT-2001.
PD
XX
XX  03-APR-2001; 2001WO-US10698.
PF
XX
XX  03-APR-2000; 2000US-194160P.
PR
XX  11-AUG-2000; 2000US-224604P.
PR
XX  17-NOV-2000; 2000US-249548P.
XX

```


PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX P-PSDB; ABB63556.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Claim 1; SEQ ID NO 17459; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CC sequences (ABLI1840-ABLI16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 2016 BP; 469 A; 511 C; 547 G; 489 T; 0 other;
SQ Query Match 47.7%; Score 24.8; DB 23; Length 2016;
Best Local Similarity 72.7%; Pred. No. 52;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 9 CTCTCAGCTCTCATCGCTCTCGTCTCGCGGCTCGCCAGTG 52
Db 590 CCCTCAGCTTCCATGCCCTCGTCGATGACTTCTCGGCATTG 547
RESULT 6
ABL07658
ID ABL07658 standard; cDNA; 4095 BP.
XX ABL07658;
AC ABL07658;
XX 26-MAR-2001; 2001WO-US09231
DT Drosophila melanogaster expressed polynucleotide SEQ ID NO 17456.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX Drosophila melanogaster.
OS Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
PF 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
PR (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX P-PSDB; ABB63555.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX

PS Claim 1; SEQ ID NO 17456; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CC sequences (ABLI1840-ABLI16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 4095 BP; 1151 A; 937 C; 945 G; 1062 T; 0 other;
SQ Query Match 47.7%; Score 24.8; DB 23; Length 4095;
Best Local Similarity 72.7%; Pred. No. 57;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 9 CTCTCAGCTCTCATCGCTCTCGTCTCGCGGCTCGCCAGTG 52
Db 2427 CCCTCAGCTTCCATGCCCTCGTCGATGACTTCTCGGCATTG 2470
RESULT 7
AAS57171/c
ID AAS57171 standard; cDNA; 1566 BP.
XX AAS57171;
AC AAS57171;
XX 16-JAN-2002 (first entry)
DT cDNA encoding Drosophila G-protein coupled receptor, GPCR #50.
XX Drosophila; G-protein coupled receptor; GPCR; insecticide; diagnostic;
XX mutation detection; ss.
XX Drosophila melanogaster.
OS Drosophila melanogaster.
XX WO200170980-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09341.
PF 23-MAR-2000; 2000US-191638P.
XX 18-JUL-2000; 2000US-0618893.
PR (PEKE) PE CORP NY.
XX Cravchik A;
PI WPI; 2001-616405/71.
XX P-PSDB; AAU38972.
XX Sixty six Drosophila Melanogaster G-protein coupled receptors (GPCR),
PT useful in the treatment and diagnosis of GPCR-related conditions and
PT for identifying GPCR modulators for use as insecticides -
XX Claim 4; Page 229-230; 392pp; English.
XX The invention relates to sixty six novel isolated Drosophila
CC melanogaster G-protein coupled receptors (GPCR). The GPCR proteins and
CC nucleic acids are useful in the treatment and diagnosis of GPCR-related
CC conditions. The GPCR proteins and nucleic acids are also useful for
CC identifying modulators of GPCR proteins for use as insecticides. The
CC nucleic acid can also be used to detect mutations in GPCR genes and gene
CC expression products such as mRNA. AAS57072-AAS57203 represent D.
CC melanogaster G-coupled protein receptor genomic and coding sequences of
CC the invention.
XX Sequence 1566 BP; 291 A; 490 C; 437 G; 348 T; 0 other;
SQ

```
Query Match      47.3%; Score 24.6; DB 22; Length 1566;
Best Local Similarity 70.2%; Pred. No. 59;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      3 GCAGTTCTCTCAGGCTCTCATCGCTCTCGTCTCGTCCGCGCTCGCCA 49
DB      646 GCACAGCGGTACGACCTCTCGCTCTCTCTCGTCTCGTGGAGTCTCCACA 600

RESULT 8
AAH49426/c
ID   AAH49426 standard; cDNA; 1566 BP.
XX
AC   AAH49426;
XX
DT   11-DEC-2001 (first entry)
XX
DE   D. melanogaster peptide receptor encoding cDNA SEQ ID 43.
XX
KW   Insect; fruitfly; peptide receptor; plant protection; insecticide; ss.
XX
OS   Drosophila melanogaster.
XX
PN   DE10013618-Al.
XX
PD   20-SEP-2001.
XX
PF   18-MAR-2000; 2000DE-1013618.
XX
PR   18-MAR-2000; 2000DE-1013618.
XX
PA   (FARB ) BAYER AG.
XX
PI   Antonicek H, Friedrich G, Schulte T;
XX
WPI; 2001-571695/65.
DR   P-PSDB; AAB86973.
XX
New polypeptides from Drosophila melanogaster have biological activity
PT of peptide receptor, useful to find new compounds for plant protection
PT and insecticides -
XX
PS   Claim 6; Page 107-110; 128pp; German.
XX
This invention describes novel polypeptides (p1) from Drosophila
CC melanogaster having the biological activity of a peptide receptor.
CC Molecules of the invention are used to find new plant protection
CC compounds of insecticides, or to find genes encoding a polypeptide
CC involved in the structure of functionally similar receptors in insects
CC This sequence encodes a Drosophila melanogaster (fruitfly) peptide
CC receptor described in the method of the invention.
XX
SQ   Sequence 1566 BP; 291 A; 490 C; 437 G; 348 T; 0 other;

Query Match      47.3%; Score 24.6; DB 22; Length 1566;
Best Local Similarity 70.2%; Pred. No. 59;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      3 GCAGTTCTCTCAGGCTCTCATCGCTCTCGTCTCGTCCGCGCTCGCCA 49
DB      646 GCACAGCGGTACGACCTCTCGCTCTCTCTCGTCTCGTGGAGTCTCCACA 600

RESULT 9
ABL22999/c
ID   ABL22999 standard; DNA; 1566 BP.
XX
AC   ABL22999;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 20470.

Query Match      47.3%; Score 24.6; DB 22; Length 1566;
Best Local Similarity 70.2%; Pred. No. 59;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      3 GCAGTTCTCTCAGGCTCTCATCGCTCTCGTCTCGTCCGCGCTCGCCA 49
DB      646 GCACAGCGGTACGACCTCTCGCTCTCTCTCGTCTCGTGGAGTCTCCACA 600

RESULT 10
AAS57170/c
ID   AAS57170 standard; DNA; 3772 BP.
XX
AC   AAS57170;
XX
DT   16-JAN-2002 (first entry)
XX
DE   DNA encoding Drosophila G-protein coupled receptor, GPCR #50.
XX
KW   Drosophila; G-protein coupled receptor; GPCR; insecticide; diagnostic;
mutation detection; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200170980-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09341.
XX
PR   23-MAR-2000; 2000US-191638P.
PR   18-JUL-2000; 2000US-0618893.
```

```
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 20470; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1566 BP; 291 A; 490 C; 437 G; 348 T; 0 other;

Query Match      47.3%; Score 24.6; DB 23; Length 1566;
Best Local Similarity 70.2%; Pred. No. 59;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      3 GCAGTTCTCTCAGGCTCTCATCGCTCTCGTCTCGTCCGCGCTCGCCA 49
DB      646 GCACAGCGGTACGACCTCTCGCTCTCTCTCGTCTCGTGGAGTCTCCACA 600

RESULT 10
AAS57170/c
ID   AAS57170 standard; DNA; 3772 BP.
XX
AC   AAS57170;
XX
DT   16-JAN-2002 (first entry)
XX
DE   DNA encoding Drosophila G-protein coupled receptor, GPCR #50.
XX
KW   Drosophila; G-protein coupled receptor; GPCR; insecticide; diagnostic;
mutation detection; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200170980-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09341.
XX
PR   23-MAR-2000; 2000US-191638P.
PR   18-JUL-2000; 2000US-0618893.
```


CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15137 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.

XX Sequence 695 BP; 144 A; 199 C; 174 G; 174 T; 4 other;

SQ Query Match 46.9%; Score 24.4; DB 21; Length 695;

Best Local Similarity 73.8%; Pred. No. 63;

Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 6 GTTCTCTACGCTCTCATCGCTCTCGTCTCGTGGCGCTCGC 47

Db 186 GATCTCTTACGCGCTCATCGATGACATCGTCGACGCGCTTGC 227

RESULT 13

AAST3006/c

ID AAST3006 standard; cDNA; 5484 BP.

XX

AC AAST3006;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #8810.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

XX 31-MAR-2000; 2000US-0540217.

PR

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

XX WPI; 2001-639362/73.

DR

DR P-PSDB; ABG08819.

XX

XX Claim 1; SEQ ID No 8810; 103pp; English.

PS

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5484 BP; 1153 A; 1384 C; 1583 G; 1364 T; 0 other;

SQ Query Match 46.9%; Score 24.4; DB 23; Length 5484;

Best Local Similarity 73.8%; Pred. No. 80;

Matches 31; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 7 TTCTCTACGCTCTCATCGCTCTCGTCTCGTGGCGCTCGCC 48

Db 4438 TTGCTCGCTCTCTCGTCTCGTCTCGTCTCGCTCGCGCGCTCGCC 4397

RESULT 14

AAK70862

ID AAK70862 standard; DNA; 25913 BP.

XX

AC AAK70862;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25674.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KM cytostatic; gene therapy; vaccine; metastasis; ds.

XX

OS Homo sapiens.

XX

PN WO200157182-A2.

XX

PD 09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01354.

XX

XX 31-JAN-2000; 2000US-0179065.

PR

PR 04-FEB-2000; 2000US-0180628.

PR

PR 24-FEB-2000; 2000US-0184664.

PR

PR 02-MAR-2000; 2000US-0186350.

PR

PR 16-MAR-2000; 2000US-0189874.

PR

PR 17-MAR-2000; 2000US-0190076.

PR

PR 18-APR-2000; 2000US-0198123.

PR

PR 19-MAY-2000; 2000US-0205515.

PR

PR 07-JUN-2000; 2000US-0209467.

PR

PR 30-JUN-2000; 2000US-0215135.

PR

PR 07-JUL-2000; 2000US-0216647.

PR

PR 07-JUL-2000; 2000US-0216880.

PR

PR 11-JUL-2000; 2000US-0217487.

PR

PR 14-JUL-2000; 2000US-0217496.

PR

PR 26-JUL-2000; 2000US-0220963.

PR

PR 26-JUL-2000; 2000US-0220964.

PR

PR 14-AUG-2000; 2000US-0224518.

PR

PR 14-AUG-2000; 2000US-0224519.

PR

PR 14-AUG-2000; 2000US-0225213.

PR

PR 14-AUG-2000; 2000US-0225214.

PR

PR 14-AUG-2000; 2000US-0225266.

PR

PR 14-AUG-2000; 2000US-0225267.

PR

PR 14-AUG-2000; 2000US-0225268.

PR

PR 14-AUG-2000; 2000US-0225270.

PR

PR 14-AUG-2000; 2000US-0225447.

PR

PR 14-AUG-2000; 2000US-0225757.

PR

PR 14-AUG-2000; 2000US-0225758.

PR

PR 14-AUG-2000; 2000US-0225759.

PR

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02595678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 25674; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 25913 BP; 5398 A; 6705 C; 7089 G; 6721 T; 0 other;

Query Match 46.9%; Score 24.4; DB 22; Length 25913;
Best Local Similarity 68.0%; Pred. No. 96;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
Qy 2 TGCAGTTCTCTCACGCTCTCATCGCTCGCTCGCTCGCGCTCGCCAGT 51
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1099 TGCAGTTCCCTCACAGTCTTCACGTTCTCTTCCTTCGGAAGCTCCAGT 1148

RESULT 15
ABK83564/c
ID ABK83564 standard; cDNA; 198161 BP.
XX AC ABK83564;
XX
XX 14-AUG-2002 (first entry)
XX Human cDNA differentially expressed in granulocytic cells #135.
DE
XX
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
XX Homo sapiens.
XX
XX WO200228999-A2.
XX
XX 11-APR-2002.
XX
XX 03-OCT-2001; 2001WO-US30821.
XX
XX 03-OCT-2000; 2000US-237189P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX
XX
XX Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
XX Claim 1; SEQ ID No 135; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is

CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 198161 BP; 61561 A; 45946 C; 41360 G; 49294 T; 0 other;
SQ
Query Match 46.9%; Score 24.4; DB 24; Length 198161;
Best Local Similarity 68.0%; Pred. No. 1.2e+02;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
OY 2 TGCAGTTCCTCAGCGCTCTCATCGCTCTCGTCGCTCGCGCCTCGCCAGT 51
Db 25296 TGCAGTTCCTCAGCGCTCTCATCGCTCTCTTCCTTCGGAAGCTCCAGT 25247
RESULT 16
AAA94724
ID AAA94724 standard; DNA; 2162 BP.
XX
XX AAA94724;
AC
XX 02-FEB-2001 (first entry)
XX
XX Rice auxin transport protein clone rsl1n.pk003.n3 DNA sequence.
DE
XX
XX Auxin transport protein; rice;
KW root development; gene mapping; plant breeding; herbicide; ss.
XX
XX Oryza sativa.
XX
XX WO200068389-A2.
XX
XX 16-NOV-2000.
XX
XX 03-MAY-2000; 2000WO-US12061.
XX
XX 07-MAY-1999; 99US-0133040.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX Orozco EM, Weng Z, Bruce WB, Cahoon RE, Tao Y;
XX WPI; 2000-687647/67.
XX P-PSDB; AAB26937.
XX
XX New nucleic acid sequences encoding new auxin transport proteins,
PT useful for modulating root growth of plants and to screen for
PT herbicides -
XX
XX Claim 2; Pages 65-66; 94pp; English.
XX
XX Auxins are plant hormones that influence plant behaviour and development
CC e.g. vascular tissue differentiation, apical development, tropic
CC responses and organ (e.g. flower, leaf) development. The present
CC invention relates to corn auxin transport protein coding sequences and
CC proteins. The present sequence is one such coding sequence. This sequence
CC may be used to modulate root development, e.g. to produce a more robust
CC root system, alter root angle or redirect root growth. Also, the present
CC sequence may be useful for gene mapping (e.g. for plant breeding) and to
CC identify loss of function mutants. The protein encoded by the present
CC sequence may be useful for raising specific antibodies, for the detection
CC of auxin transport proteins and to design and/or identify specific
CC inhibitors of auxin transport proteins, potentially useful as herbicides.
XX
XX Sequence 2162 BP; 428 A; 694 C; 636 G; 404 T; 0 other;
SQ
Query Match 46.2%; Score 24; DB 21; Length 2162;
Best Local Similarity 68.8%; Pred. No. 97;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can
CC be modulated. The presence of (1) or HA proteins encoded by then are
CC used for diagnosing the presence or activity of *Corynebacterium*
CC diphtheriae. (1) can be used to map the *C. glutamicum* genome or can be
CC used as markers for genetically engineered *Corynebacterium* or
CC *Brevibacterium*. The HA proteins encoded by the (1) are used to maintain
CC homeostasis in *C. glutamicum* or help the microorganism to adapt to
CC different environmental conditions.
XX
SQ Sequence 732 BP; 183 A; 220 C; 166 G; 163 T; 0 other;

Query Match 45.8%; Score 23.8; DB 22; Length 732;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 34; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 TGCAGTCTCTCAGGCTCTCATCGCTCTCGTCTCGCTCGCGGCTCGCCAGTG 52
DB 501 TTCAGATCGGATACGCCCATCATCGGTATCCGCGCAGCAGGTCTCGCGCTG 551

RESULT 19
AB259094/C
ID AB259094 standard; cDNA; 7077 BP.
XX AC AB259094;
XX
XX 28-APR-2003 (first entry)
DT
DE E. maxima 230 kDa gametocyte antigen cDNA.
XX
XX Gametocyte; antigen; 56 kDa; 82 kDa; antiparasitic; vaccine;
KW 230 kDa; Gene; ss.
XX
XX Eimeria maxima.
OS
XX
PN WO2003004683-A2.
XX
XX 16-JAN-2003.
PD
XX
PF 03-JUL-2002; 2002WO-US21233.
XX
PR 06-JUL-2001; 2001US-303699P.
XX
XX (BELL/) BELLI S I.
PA (SMIT/) SMITH N C.
PA (WALL/) WALLACH M.
XX
XX Belli SI, Smith NC, Wallach M;
PI
XX
XX WPI; 2003-221605/21.
DR
XX
XX New nucleic acid comprising a sequence encoding a 56 kDa or 82 kDa
PT polypeptide from gametocytes of *Eimeria* maxima, useful for preparing a
PT vaccine against *Eimeria* infection -
XX
XX Disclosure; Fig 12; 112pp; English.
PS
XX
XX The invention provides two major E. maxima gametocyte antigens having
CC molecular weights of 56 and 82 kDa and encoding polynucleotides. The
CC gametocyte antigens can be produced by standard recombinant methodology.
CC The encoding polynucleotides are useful for preparing a vaccine against
CC *Eimeria* tenella, *E. acervulina*, *E. necatrix*, *E. praecox*, *E. maxima*,
CC *E. mitis* or *E. brunetti* infection. The present sequence represents a
CC E. maxima 230 kDa gametocyte antigen cDNA clone.
XX
SQ Sequence 7077 BP; 1814 A; 1987 C; 2163 G; 1113 T; 0 other;

Query Match 45.8%; Score 23.8; DB 25; Length 7077;
Best Local Similarity 72.1%; Pred. No. 1.3e+02;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 7 TTCTCTCAGGCTCTCATCGCTCTCGTCTCGTCTCGCGGCTCGCCCA 49

DB 2781 TTCTTCTGCGCACTCATCAATGCTTGGCAGCGCTCGCTCGCCA 2739

RESULT 20
AB222977/C
ID AB222977 standard; DNA; 7077 BP.
XX AC AB222977;
XX
XX 10-APR-2003 (first entry)
DT
DE Eimeria maxima 250 kDa protein homologous DNA sequence SEQ ID NO:4.
XX
XX Eimeria maxima; 250 kDa antigen; sporozoite; merozoite; infection;
KW vaccine; antiparasitic; gene therapy; immunodominant; gene; ds.
XX
XX Eimeria maxima.
OS
XX Synthetic.
XX
PN WO2003004684-A2.
XX
XX 16-JAN-2003.
PD
XX
PF 03-JUL-2002; 2002WO-US21237.
XX
PR 06-JUL-2001; 2001US-303670P.
XX
XX (WITC/) WITCOMBE D.
PA (SMIT/) SMITH N C.
PA (WALL/) WALLACH M.
XX
XX Witcombe D, Smith NC, Wallach M;
PI
XX
XX WPI; 2003-201556/19.
DR
XX
XX New nucleic acid comprising a sequence encoding a 250 kDa polypeptide
PT from sporozoites/merozoites of *Eimeria* maxima, useful for preparing a
PT vaccine against *Eimeria* infection -
XX
XX Claim 15; Fig 7A; 198pp; English.
PS
XX
XX The present invention describes a nucleic acid (1) comprising a
CC sequence encoding a 250 kDa polypeptide from Sporozoites/Merozoites of
CC *Eimeria* maxima or its homologue or complement. Also described: (1) a
CC vector comprising the nucleic acid; (2) a host cell comprising the
CC vector; (3) a plasmid comprising the nucleic acid; (4) a transformed
CC cell comprising the nucleic acid; (5) producing a recombinant 250 kDa
CC polypeptide; (6) a recombinant polypeptide; (7) a vaccine against
CC *E. tenella*, *E. acervulina*, *E. necatrix*, *E. praecox*, *E. maxima*, *E. mitis*
CC or *E. brunetti* or a microorganism expressing an immunologically
CC cross-reactive antigen; (8) immunising a subject against infection by
CC *Eimeria*, or a microorganism expressing an immunologically cross-reactive
CC antigen; (9) conferring upon a newborn subject of an avian species
CC maternal immunity against infection by *Eimeria*; (10) a fertilised egg
CC from an avian species having an air sac, where the air sac is inoculated
CC with the vaccine; and (11) reducing the output of *Eimeria* oocytes in
CC faeces from a newborn subject of an avian species. (1) has antiparasitic
CC activity and can be used in a vaccine and in gene therapy. The nucleic
CC acid is useful for preparing a vaccine against *E. tenella*, *E. acervulina*,
CC *E. necatrix*, *E. praecox*, *E. maxima*, *E. mitis* or *E. brunetti* infection.
CC The present sequence represents a 250 kDa antigen protein homologous DNA
CC sequence, which is used in an example from the present invention.
XX
SQ Sequence 7077 BP; 1814 A; 1985 C; 2165 G; 1113 T; 0 other;

Query Match 45.8%; Score 23.8; DB 25; Length 7077;
Best Local Similarity 72.1%; Pred. No. 1.3e+02;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 7 TTCTCTCAGGCTCTCATCGCTCTCGTCTCGTCTCGCGGCTCGCCCA 49
DB 2781 TTCTTCTGCGCACTCATCAATGCTTGGCAGCGCTCGCTCGCCA 2739

RESULT 21
AB222976/c
ID AB222976 standard; cDNA; 7987 BP.
XX AC AB222976;
XX DT 10-APR-2003 (first entry)
XX DE Eimeria maxima 250 kDa protein encoding cDNA SEQ ID NO:1.
XX KW Eimeria maxima; 250 kDa antigen; sporozoite; merozoite; infection;
XX KW vaccine; antiparasitic; gene therapy; immunodominant; gene; ss.
XX OS Eimeria maxima.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT CDS 3..7985
FT FT /*tag= a
FT FT /partial
FT FT /product= "250 kDa antigen protein"
FT FT /note= "the coding sequence is given as shown in the
FT FT figure and any X's in the protein sequence are encoded
FT FT by stop codons; no start or stop codons given"
XX PN WO2003004684-A2.
XX XX
XX PD 16-JAN-2003.
XX PF 03-JUL-2002; 2002WO-US21237.
XX PR 06-JUL-2001; 2001US-303670P.
XX PA (WITC/) WITCOMBE D.
XX PA (SMIT/) SMITH N C.
XX PA (WALL/) WALLACH M.
XX PI Witcombe D, Smith NC, Wallach M;
XX WPI; 2003-201556/19.
XX DR P-PSDB; ABP56958.
XX XX
XX PT New nucleic acid comprising a sequence encoding a 250 kDa polypeptide
XX PT from sporozoites/merozoites of Eimeria maxima, useful for preparing a
XX PT vaccine against Eimeria infection -
XX PS Example 3; Fig 6; 198pp; English.
XX CC The present invention describes a nucleic acid (I) comprising a
XX CC sequence encoding a 250 kDa polypeptide from Sporozoites/Merozoites of
XX CC Eimeria maxima or its homologue or complement. Also described: (1) a
XX CC vector comprising the nucleic acid; (2) a host cell comprising the
XX CC cell comprising the nucleic acid; (5) producing a recombinant 250 kDa
XX CC polypeptide; (6) a recombinant polypeptide; (7) a vaccine against
XX CC E. tenella, E. acervulina, E. necatrix, E. praecox, E. maxima, E. mitis
XX CC or E. brunetti or a microorganism expressing an immunologically
XX CC cross-reactive antigen; (8) immunising a subject against infection by
XX CC Eimeria, or a microorganism expressing an immunologically cross-reactive
XX CC antigen; (9) conferring upon a newborn subject of an avian species
XX CC maternal immunity against infection by Eimeria; (10) a fertilised egg
XX CC from an avian species having an air sac, where the air sac is inoculated
XX CC with the vaccine; and (11) reducing the output of Eimeria oocytes in
XX CC faeces from a newborn subject of an avian species. (I) has antiparasitic
XX CC activity and can be used in a vaccine and in gene therapy. The nucleic
XX CC acid is useful for preparing a vaccine against E. tenella, E. acervulina,
XX CC E. necatrix, E. praecox, E. maxima, E. mitis or E. brunetti infection.
XX CC The present sequence encodes a 250 kDa antigen protein, which is used
XX CC in an example from the present invention.
XX SQ Sequence 7987 BP; 2022 A; 2217 C; 2370 G; 1378 T; 0 other;

Query Match 45.8%; Score 23.8; DB 25; Length 7987;
Best Local Similarity 72.1%; Pred. No. 1.3e+02;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 7 TTCTCTCAGCTCTCATCGCTCTCGTGGCTGGCGGCTCGCCA 49
DB 3011 TTCTTCTGGCAGCTCATCAATGCTTGGCAGCTCGCTCGCCA 2969
RESULT 22
AB222984/c
ID AB222984 standard; cDNA; 7990 BP.
XX AC AB222984;
XX DT 10-APR-2003 (first entry)
XX DE E. maxima immunodominant protein encoding cDNA SEQ ID NO:25.
XX KW Eimeria maxima; 250 kDa antigen; sporozoite; merozoite; infection;
XX KW vaccine; antiparasitic; gene therapy; gene; ss.
XX OS Eimeria maxima.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT CDS 231..7313
FT FT /*tag= a
FT FT /product= "immunodominant protein"
FT FT sig_peptide 231..308
FT FT /*tag= b
FT FT mat_peptide 309..7310
FT FT /*tag= c
XX PN WO2003004684-A2.
XX XX
XX PD 16-JAN-2003.
XX PF 03-JUL-2002; 2002WO-US21237.
XX PR 06-JUL-2001; 2001US-303670P.
XX PA (WITC/) WITCOMBE D.
XX PA (SMIT/) SMITH N C.
XX PA (WALL/) WALLACH M.
XX PI Witcombe D, Smith NC, Wallach M;
XX WPI; 2003-201556/19.
XX DR P-PSDB; ABP56972.
XX XX
XX PT New nucleic acid comprising a sequence encoding a 250 kDa polypeptide
XX PT from sporozoites/merozoites of Eimeria maxima, useful for preparing a
XX PT vaccine against Eimeria infection -
XX PS Example 5; Fig 29; 198pp; English.
XX CC The present invention describes a nucleic acid (I) comprising a
XX CC sequence encoding a 250 kDa polypeptide from Sporozoites/Merozoites of
XX CC Eimeria maxima or its homologue or complement. Also described: (1) a
XX CC vector comprising the nucleic acid; (2) a host cell comprising the
XX CC cell comprising the nucleic acid; (5) producing a recombinant 250 kDa
XX CC polypeptide; (6) a recombinant polypeptide; (7) a vaccine against
XX CC E. tenella, E. acervulina, E. necatrix, E. praecox, E. maxima, E. mitis
XX CC or E. brunetti or a microorganism expressing an immunologically
XX CC cross-reactive antigen; (8) immunising a subject against infection by
XX CC Eimeria, or a microorganism expressing an immunologically cross-reactive
XX CC antigen; (9) conferring upon a newborn subject of an avian species
XX CC maternal immunity against infection by Eimeria; (10) a fertilised egg
XX CC from an avian species having an air sac, where the air sac is inoculated
XX CC with the vaccine; and (11) reducing the output of Eimeria oocytes in
XX CC faeces from a newborn subject of an avian species. (I) has antiparasitic
XX CC activity and can be used in a vaccine and in gene therapy. The nucleic
XX CC acid is useful for preparing a vaccine against E. tenella, E. acervulina,
XX CC E. necatrix, E. praecox, E. maxima, E. mitis or E. brunetti infection.
XX CC The present sequence encodes a 250 kDa antigen protein, which is used
XX CC in an example from the present invention.


```

XX OS Streptomyces verticillus.
XX FH Key
XX CDS Location/Qualifiers
FT     223..564
FT     /tag= a
FT     /transl_except= (pos: 1..3, aa: Met)
FT     /note= "ORF 30; encodes AAB07556"
FT     561..2309
FT     /tag= b
FT     /transl_except= (pos: 1..3, aa: Met)
FT     /note= "ORF 29; encodes AAB07557"
FT     2767..3486
FT     /tag= c
FT     /note= "ORF 28; encodes AAB07558"
FT     3527..5593
FT     /tag= d
FT     /transl_except= (pos: 1..3, aa: Met)
FT     /note= "ORF 27; encodes AAB07559"
FT     5806..12294
FT     /tag= e
FT     /note= "ORF 26; encodes AAB07560"
FT     12291..15491
FT     /tag= f
FT     /note= "ORF 25; encodes AAB07561"
FT     15488..21013
FT     /tag= g
FT     /note= "ORF 24; encodes AAB07562"
FT     21010..24666
FT     /tag= h
FT     /transl_except= (pos: 1..3, aa: Met)
FT     /note= "ORF 23; encodes AAB07563"
FT     24663..32690
FT     /tag= i
FT     /note= "ORF 22; encodes AAB07564"
FT     32893..34830
FT     /tag= j
FT     /note= "ORF 21; encodes AAB07565"
FT     34827..35804
FT     /tag= k
FT     /transl_except= (pos: 1..3, aa: Met)
FT     /note= "ORF 20; encodes AAB07566"
FT     35818..37302
FT     /tag= l
FT     /transl_except= (pos: 1..3, aa: Met)
FT     /note= "ORF 19; encodes AAB07567"
FT     37299..39215
FT     /tag= m
FT     /transl_except= (pos: 1..3, aa: Met)
FT     /note= "ORF 18; encodes AAB07568"
FT     39301..47181
FT     /tag= n
FT     /note= "ORF 17; encodes AAB07569"
FT     47178..49985
FT     /tag= o
FT     /transl_except= (pos: 1..3, aa: Met)
FT     /note= "ORF 16; encodes AAB07570"
FT     49982..51001
FT     /tag= p
FT     /transl_except= (pos: 1..3, aa: Met)
FT     /note= "ORF 15; encodes AAB07571"
FT     50998..52386
FT     /tag= q
FT     /transl_except= (pos: 1..3, aa: Met)
FT     /note= "ORF 14; encodes AAB07572"
FT     52383..52946
FT     /tag= r
FT     /note= "ORF 13; encodes AAB07573"
FT     53018..54190
FT     /tag= s
FT     /note= "ORF 12; encodes AAB07574"
FT     54187..55824
FT     /tag= t

```

```

FT CDS /note= "ORF 11; encodes AAB07575"
FT 55821..56093
FT /*tag= u
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 10; encodes AAB07576"
FT 56090..57586
FT /*tag= v
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 9; encodes AAB07577"
FT 57583..58857
FT /*tag= w
FT /note= "ORF 8; encodes AAB07578"
FT XX
XX WO2000040704-A1.
XX 13-JUL-2000.
XX
XX 06-JAN-2000; 2000WO-US00445.
XX
XX 06-JAN-1999; 99US-0115435.
XX 05-FEB-1999; 99US-0118848.
XX 05-JAN-2000; 2000US-0477962.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Shen B, Du L, Sanchez C, Chen M, Edwards DJ;
XX
XX WPI; 2000-465974/40.
XX P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
XX AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
XX AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
XX AAB07576, AAB07577, AAB07578.
XX
XX New bleomycin gene cluster components useful for peptide and/or
XX polyketide metabolites, especially bleomycin, production and for
XX chemically modifying biological molecules -
XX
XX Claim 8; Page 97-136; 162pp; English.
XX
XX The present sequence represents the BLM (Bleomycin) gene cluster,
XX containing open reading frames (ORFs) 8-30. The proteins encoded
XX by the gene cluster are useful for producing peptides and/or polyketide
XX metabolites, especially bleomycin or bleomycin analogues. They are
XX also useful for chemically modifying biological molecules to produce
XX branched methyl groups, and for coupling amino acids and fatty
XX acids. They may be reacted with an apo-carrier protein and coenzyme A
XX to produce a holo-carrier protein. The BLM gene cluster or catalytic
XX domains can be used individually or collectively to produce
XX thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
XX microbial metabolites. The BLM gene cluster may also be used to produce
XX sugars.
XX
XX Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 other;
XX
XX Query Match 45.8%; Score 23.8; DB 21; Length 58857;
XX Best Local Similarity 66.7%; Pred. No. 1.7e+02;
XX Matches 34; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
XX
XX QY 2 TGCAGTTCTCTCACGCTCTCATGCTCTCGTCTCGCGCCCTCGCCAGTG 52
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 10310 TGCAGTTCTCTCCGACGCCCTCATGAGAGGCTCCGACAGCTAGCCCGCG 10360
XX
XX RESULT 25
XX AAH68534
XX ID AAH68534 standard; DNA; 309400 BP.
XX
XX AC AAH68534;
XX
XX DT 26-SEP-2001 (first entry)
XX
XX DE C glutamicum coding sequence fragment SEQ ID NO: 7069.
XX

```

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
OS Corynebacterium glutamicum.
XX
XX
XX EPI108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
PI
XX WPI; 2001-376931/40.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Disclosure; SEQ ID NO: 7069; 246pp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 309400 BP; 70133 A; 86477 C; 83115 G; 69675 T; 0 other;
SQ
Query Match 45.8%; Score 23.8; DB 22; Length 309400;
Best Local Similarity 66.7%; Pred. No. 2e-02;
Matches 34; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 2 TGCAGTTCTCTCAGCTCTCATCGCTCTCGCTCGCTCGCGCTCGCGCTCG 52
DB 218393 TTCAGATCGGATACGCCATCATCGGTATCGCGCAGCAGGTCTCGCGCTG 218443
RESULT 26
ABZ11723
ID ABZ11723 standard; cDNA; 1415 BP.
XX
XX ABZ11723;
XX
XX 20-JAN-2003 (first entry)
XX
XX Human polynucleotide SEQ ID NO 605.
DE
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW antiarthritic; gene; ss.
XX

OS Homo sapiens.
XX
XX WQ200270539-A2.
XX
XX 12-SEP-2002.
XX
XX 05-MAR-2002; 2002WO-US05095.
XX
XX 05-MAR-2001; 2001US-0799451.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX
XX WPI; 2002-759812/82.
DR P-PSDB; ABP69506.
XX
XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative, or
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT platelet or coagulation disorders
XX
XX Claim 1: SEQ ID NO 605; 1012pp + Sequence Listing; English.
XX
XX The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences
CC (ABZ1119-ABZ12066) or their mature protein coding portion, active domain
CC coding protein or complementary sequences. The polynucleotides are useful
CC for identifying expressed genes or for physical mapping of human genome.
CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
CC weight markers, as a food supplement, for generating antibodies, in
CC medical imaging, screening and diagnostic assays and for treating
CC cell-proliferative disorders (cancer), neurodegenerative diseases
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC disorders, platelet or coagulation disorders, wound, burns, incision,
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC parasitic), arthritis, etc.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1415 BP; 337 A; 490 C; 386 G; 202 T; 0 other;
SQ
Query Match 45.4%; Score 23.6; DB 24; Length 1415;
Best Local Similarity 69.6%; Pred. No. 1.3e+02;
Matches 32; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 ATGCAGTTCTCTCAGCTCTCATCGCTCTCGCTCGCTCGCGCTCG 46
DB 29 ATGCAGTTCTCTCAGCTCTGTGCGCCCGCTCGCTCGCGGCGCCG 74
RESULT 27
ABX34289/c
ID ABX34289 standard; DNA; 135638 BP.
XX
XX ABX34289;
XX
XX 11-FEB-2003 (first entry)
XX
XX S. atroolivaceus leinamycin biosynthesis gene cluster.
XX
XX Leinamycin biosynthesis gene cluster; Lnm; open reading frame; ORF;
KW anti-tumour antibiotic; broad spectrum antimicrobial activity;
KW Gram-positive; Gram-negative bacteria; chemical modification;
KW metabolite; apo-carrier protein; holo-carrier protein; tumour;
KW polypeptide; hybrid polypeptide/polyketide metabolite; Lnm production;
KW cytostatic; gene; ds.
XX
XX Streptomyces atroolivaceus.
OS

XX PN WO200277179-A2.
XX AC 03-OCT-2002.
XX PF 22-MAR-2002; 2002WO-US08937.
XX PR 26-MAR-2001; 2001US-278935P.
XX PA (REGC) UNIV CALIFORNIA.
XX PA (KYOW) KYOWA HAKKO KOGYO KK.
XX PI Shen B, Cheng Y, Tang G;
XX WPI; 2003-018907/01.
DR P-PSDR; ABU11341, ABU11342, ABU11343, ABU11344, ABU11345, ABU11346,
DR ABU11347, ABU11348, ABU11349, ABU11350, ABU11351, ABU11352, ABU11353,
DR ABU11354, ABU11355, ABU11356, ABU11357, ABU11358, ABU11359, ABU11360,
DR ABU11361, ABU11362, ABU11363, ABU11364, ABU11365, ABU11366, ABU11367,
DR ABU11368, ABU11369, ABU11370, ABU11371, ABU11372, ABU11373, ABU11374,
DR ABU11375, ABU11376, ABU11377, ABU11378, ABU11379, ABU11380, ABU11381,
DR ABU11382, ABU11383, ABU11384, ABU11385, ABU11386, ABU11387, ABU11388,
DR ABU11389, ABU11390, ABU11391, ABU11392, ABU11393, ABU11394, ABU11395,
DR ABU11396, ABU11397, ABU11398, ABU11399, ABU11400, ABU11401, ABU11402,
DR ABU11403, ABU11404, ABU11405, ABU11406, ABU11407, ABU11408, ABU11409,
DR ABU11410, ABU11411.
XX Novel gene cluster responsible for synthesis of leinamycin in
PT Streptomycetes atroolivaceus useful for making various peptide and/or
PT polyketide, and/or hybrid polypeptide/polyketide metabolites -
XX Claim 6; Page 81-127; 185pp; English.
XX The present invention relates to the isolation of the Streptomycetes
CC atroolivaceus leinamycin (Lnm) biosynthesis gene cluster containing
CC 71 open reading frames (ORFs) (ORFs -35 through -1, ORFs lmaA through
CC lmaZ, and ORFs +1 through +9). Leinamycin is a novel anti-tumour
CC antibiotic produced by several Streptomycetes species. It exhibits
CC broad spectrum antimicrobial activity against Gram-positive and
CC Gram-negative bacteria, but not against fungi. The polypeptides encoded
CC by the Lnm biosynthesis gene cluster ORFs are useful for chemically
CC modifying a molecule in a host cell. The host cell is a bacterium or
CC eukaryotic cell, including a mammalian, yeast, plant, fungal, or insect
CC cell. The molecule is an endogenous metabolite produced by the host
CC cell or exogenously supplied metabolite, or an amino acid, and the
CC polypeptide is a peptide synthetase or amino transferase. The
CC polypeptides encoded by the Lnm gene cluster are useful for converting
CC an apo-carrier protein to a holo-carrier protein. Lnm shows potent
CC antitumour activity in tumour models in vivo. The Lnm gene cluster
CC modules and/or catalytic domains are useful for making various peptide
CC and/or polyketide, and/or hybrid polypeptide/polyketide metabolites.
CC The proteins encoded by the ORFs are useful alone, or in combination
CC with other active domains to modify various target substrates. The
CC Lnm gene cluster is useful to upregulate endogenous Lnm production to
CC permit Lnm production in cells and/or to make various modified Lnm.
CC Lnm, its analogue, or other polyketide, peptide or hybrid
CC polyketide/peptide metabolites are useful as therapeutic agents, to
CC treat a number of disorders, depending upon the type of metabolites.
CC The present sequence represents the S. atroolivaceus leinamycin
CC biosynthesis gene cluster.
XX SQ Sequence 135638 BP; 18570 A; 49096 C; 49039 G; 18933 T; 0 other;
Query Match 45.4%; Score 23.6; DB 25; Length 135638;
Best Local Similarity 76.3%; Pred. No. 2.1e+02;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 15 CGCTCTCATCGCTCTCGTCGCGCGCCCTCGCTCGCGAGTG 52
|||||
DB 34257 CGCTCTCGCGCGCTCGCGCGCGCCCTCGCTCGCGCGTG 34220
|||||
RESULT 28

AAD38151
ID AAD38151 standard; DNA; 145 BP.
XX AC AAD38151;
XX DT 10-SEP-2002 (first entry)
XX DE pCON3 deletion mutant construct intron A fragment.
XX KW Cytomegalovirus; CMV; expression construct; gene therapy; cytostatic;
KW human cytomegalovirus; hCMV; malignant cancer; immunisation; intron A;
KW ds.
XX Unidentified.
OS WO200231137-A2.
PN 18-APR-2002.
PD 12-OCT-2001; 2001WO-US32050.
PF 13-OCT-2000; 2000US-240502P.
PR (CHIR) CHIRON CORP.
PA Thudium K, Selby M, Ulmer J;
XX WPI; 2002-452346/48.
XX Novel human cytomegalovirus Intron A fragment for use in expression
PT constructs, lacks full-length Intron A sequence, and enhance expression
PT levels when present in expression constructs -
XX Disclosure; Page 43; 44pp; English.
XX The invention relates to cytomegalovirus (CMV) intron A fragments for
CC use in expression construct for expressing gene products. Recombinant
CC expression construct is useful for producing a recombinant polypeptide.
CC Human cytomegalovirus (hCMV) intron A fragment is useful in expression
CC constructs to express a wide variety of substances including peptides.
CC It is also useful for producing proteins useful for treating a variety
CC of malignant cancers, and for producing proteins useful for prevention,
CC treatment and/or diagnosis of a wide variety of diseases. Recombinant
CC expression construct is used in nucleic acid immunisation and gene
CC therapy. The present sequence is pCON3 deletion mutant construct
CC intron A fragment used in the invention.
XX SQ Sequence 145 BP; 27 A; 46 C; 27 G; 45 T; 0 other;
Query Match 45.0%; Score 23.4; DB 24; Length 145;
Best Local Similarity 67.3%; Pred. No. 1.1e+02;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 1 ATGCGAGTCTCTCACGCTCTCATCGCTCTCGTCGTCGCGCGCTCGCCA 49
|||||
DB 25 AGGCACACCCCTTGGCTCTTATGCACTCTCGTCTCGCGCGCGCCA 73
|||||
RESULT 29
ABN20845
ID ABN20845 standard; cDNA; 450 BP.
XX AC ABN20845;
XX DT 24-JUN-2002 (first entry)
XX DE Human ORFX polynucleotide sequence SEQ ID NO:10167.
XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;

KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KW myasthenia gravis; gene; ss.
 XX Homo sapiens.
 XX WO200192523-A2.
 XX 06-DEC-2001.
 XX 29-MAY-2001; 2001WO-US10836.
 XX 30-MAY-2000; 2000US-206132P.
 XX 29-AUG-2000; 2000US-228716P.
 XX (CURA-) CURAGEN CORP.
 XX Shinkets RA, Leach MD;
 XX WPI; 2002-106308/14.
 XX P-PSDB; ABP05093.
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders -
 XX
 XX Disclosure; SEQ ID 10167; 1037pp; English.
 XX The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 450 BP; 86 A; 106 C; 123 G; 133 T; 2 other;
 SQ
 Query Match 45.0%; Score 23.4; DB 24; Length 450;
 Best Local Similarity 81.8%; Pred. No. 1.3e+02;
 Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 15 CGCTCTCATCGCTCTCGCTCGCTCGCGGCTCGC 47
 DB 145 CGCTCTCATTCGCTCGCTCGCTCGCTCGCTCGCTCGC 177
 RESULT 30
 AAA94737
 ID AAA94737 standard; DNA; 855 BP.
 XX AAA94737;
 AC
 XX 02-FEB-2001 (first entry)
 DT
 XX

DE Corn auxin transport protein cione cillc.pk001.b7 DNA sequence.
 XX
 KW Auxin transport protein; corn;
 KW root development; gene mapping; plant breeding; herbicide; ss.
 XX
 OS Zea mays.
 XX WO200068389-A2.
 XX 16-NOV-2000.
 XX 03-MAY-2000; 2000WO-US12061.
 XX 07-MAY-1999; 99US-0133040.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX (PION-) PIONEER HI-BRED INT INC.
 XX Orozco EM, Weng Z, Bruce WB, Cahoon RE, Tao Y;
 XX WPI; 2000-687647/67.
 XX P-PSDB; AAB26952.
 XX New nucleic acid sequences encoding new auxin transport proteins,
 PT useful for modulating root growth of plants and to screen for
 PT herbicides -
 XX
 XX Claim 2; Page 93; 94pp; English.
 XX Auxins are plant hormones that influence plant behaviour and development
 CC e.g. vascular tissue differentiation, apical development, tropic
 CC responses and organ (e.g. flower, leaf) development. The present
 CC invention relates to corn auxin transport protein coding sequences and
 CC proteins. The present sequence is one such coding sequence. This sequence
 CC may be used to modulate root development, e.g. to produce a more robust
 CC root system, alter root angle or redirect root growth. Also, the present
 CC sequence may be useful for gene mapping (e.g. for plant breeding) and to
 CC identify loss of function mutants. The protein encoded by the present
 CC sequence may be useful for raising specific antibodies, for the detection
 CC of auxin transport proteins and to design and/or identify specific
 CC inhibitors of auxin transport proteins, potentially useful as herbicides.
 XX
 SQ Sequence 855 BP; 152 A; 308 C; 266 G; 129 T; 0 other;
 Query Match 45.0%; Score 23.4; DB 21; Length 855;
 Best Local Similarity 73.2%; Pred. No. 1.4e+02;
 Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 12 TCACGCTCTCATCGCTCTCGCTCGCTCGCGGCTCGCGCAGTG 52
 DB 80 TCACGCTCTCTCTCCCTGTCGACGCTGCCCAACACGCTGGTG 120
 RESULT 31
 AAC55847
 ID AAC55847 standard; DNA; 2466 BP.
 XX
 AC AAC55847;
 XX
 DT 19-JAN-2001 (first entry)
 XX
 DE Mitomycin biosynthetic gene cluster related alpha amylase orf20.
 XX
 KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;
 KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW fungicide; pesticide; ds.
 XX
 OS Streptomyces lavendulae.
 XX
 XX WO200053737-A2.
 XX 14-SEP-2000.

```
XX 10-MAR-2000; 2000WO-US06394.
XX 12-MAR-1999; 99US-0266965.
XX (MINU ) UNIV MINNESOTA.
XX (SHER/) SHERMAN D H.
XX (MAOY/) MAO Y.
XX (VARO/) VAROGLU M.
XX (HEMM/) HE M.
XX (SHEL/) SHELDON P C.
XX
XX Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
XX
XX WPI; 2000-601980/57.
XX
XX Novel nucleic acid molecule comprising mitomycin biosynthetic gene
XX cluster useful for cloning mitomycin biosynthetic genes for elucidating
XX the molecular basis of mitosane ring system biosynthesis -
XX
XX Example 1; Page 324-325; 399pp; English.
XX
XX This invention relates to isolated and purified nucleic acid molecules
XX from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
XX natural products that contain a variety of functional groups, including
XX amino benzquinone and axiridine ring systems. The S. lavendulae
XX mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
XX spanning 55kb of DNA. The invention includes an expression cassette
XX comprising a mitomycin biosynthetic gene operably linked to a promoter,
XX and host cells transformed with the cassette. The nucleotide, and protein
XX sequences and the transformed host cells of the invention result in
XX antitastmatic, antiinflammatory, cytostatic, immunomodulatory, and
XX antibiotic activities. The nucleotide sequences are used to elucidate the
XX molecular basis for the biosynthesis of the mitosane ring system, as well
XX as to engineer the biosynthesis of novel natural products, e.g.
XX antibiotics, anti-inflammatory agents, anti-cancer agents,
XX immune-enhancers, immunosuppressants, agents to treat asthma, chronic
XX obstructive pulmonary disease as well as other disease involving
XX respiratory inflammation, or cholesterol-lowering agents or as crop
XX protection agents (e.g. fungicides or insecticides) as well as
XX biopolymers, e.g., in packaging or biomedical applications, or to engineer
XX PHA monomer syntheses. Sequences AAC55782-C55881, AAC55815-C55849 and
XX AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
XX sequences and encoded proteins. Sequences AAC55812-C55814,
XX AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
XX cloning of the mitomycin biosynthetic genes.
XX
XX Sequence 2466 BP; 460 A; 860 C; 822 G; 324 T; 0 other;
XX
XX Query Match 45.0%; Score 23.4; DB 21; Length 2466;
XX Best Local Similarity 73.2%; Pred. No. 1.6e+02;
XX Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
XX 9 CTCTCACGCTCTCATCGCTCTCGTCCGCTCGCGGCTCGGCA 49
XX ||||| ||||| ||||| ||||| |||||
XX 48 CGCTGCGGCGCTCATGCGCTCTCGTCCGCTCGCGCACCGCA 88
XX
XX
XX RESULT 32
XX ABL13923
XX ID ABL13923 standard; cDNA; 2840 BP.
XX
XX ABL13923;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 36251.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 19811.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
```

```
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX P-PSDB; ABB69820.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 36251; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX sequences (ABLI01840-ABLI6175) and the encoded proteins
XX (ABB5737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2840 BP; 671 A; 724 C; 723 G; 722 T; 0 other;
XX
XX Query Match 45.0%; Score 23.4; DB 23; Length 2840;
XX Best Local Similarity 73.2%; Pred. No. 1.6e+02;
XX Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
XX 10 TCTCAGCTCTCATCGCTCTCGTCCGCTCGCGGCTCGCCAG 50
XX ||||| ||||| ||||| ||||| |||||
XX 2774 TCCTCCCGATCTCTGGTGTCTCTCGTCCGCGCTCGCCCG 2814
XX
XX
XX RESULT 33
XX ABL08443/c
XX ID ABL08443 standard; cDNA; 5736 BP.
XX
XX ABL08443;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 19811.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
```

DR WPI; 2001-656860/75.
DR P-PSDB; ABB64340.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
PS Claim 1; SEQ ID NO 19811; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLL16176-ABLL30511), expressed DNA
CC sequences (ABLL01840-ABLL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5736 BP; 1503 A; 1718 C; 1588 G; 927 T; 0 other;
Query Match 45.0%; Score 23.4; DB 23; Length 5736;
Best Local Similarity 67.3%; Pred. No. 1.7e+02;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
XX
QY 4 CAGTTCTCTCAGGCTCTCATCGCTCTCGTCTCGGCTCGGCTCGGCTCGGCTG 52
DB 5544 CAGTTCTCTCAGGCTCTCATCGCTCTCGTCTCGGCTCGGCTCGGCTCGGCTG 5496
RESULT 34
ABL13922
ID ABL13922 standard; cDNA; 6782 BP.
XX
AC ABL13922;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36248.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR P-PSDB; ABB69819.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 36248; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLL16176-ABLL30511), expressed DNA
CC sequences (ABLL01840-ABLL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6782 BP; 1899 A; 1548 C; 1559 G; 1776 T; 0 other;
Query Match 45.0%; Score 23.4; DB 23; Length 6782;
Best Local Similarity 73.2%; Pred. No. 1.8e+02;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
QY 10 TCTCAGCTCTCATCGCTCTCGTCTCGGCTCGGCTCGGCTCGGCTG 50
DB 5716 TCCCTCCGATCCTCGGTGTCCTCGTCTCGGCTCGGCTCGGCTG 5756
RESULT 35
ABL08442/c
ID ABL08442 standard; cDNA; 9577 BP.
XX
AC ABL08442;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 19808.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR P-PSDB; ABB64339.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 19808; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLL16176-ABLL30511), expressed DNA
CC sequences (ABLL01840-ABLL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 9577 BP; 2647 A; 2491 C; 2387 G; 2052 T; 0 other;
Query Match 45.0%; Score 23.4; DB 23; Length 9577;
Best Local Similarity 67.3%; Pred. No. 1.8e+02;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 4 CAGTTCTCAGGCTCTCATCGCTCTCGTGGTCCGCGCTCGCCAGTG 52
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8249 CAGTTCTCAGGCTCTCGTGGTCCCTCTCGTGGTCTCATCGTCTCGTTG 8201

RESULT 36
 AAC55841
 ID AAC55841 standard; DNA; 18034 BP.
 XX
 AC AAC55841;
 XX
 DT 19 JAN-2001 (first entry)
 XX
 DE Complete Mitomycin ORF 11-22 nucleotide sequence.
 XX
 KW Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;
 KW anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW fungicide; pesticide; ds.
 XX
 OS Streptomyces lavendulae.
 XX
 PN WO200053737-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 10-MAR-2000; 2000WO-US06394.
 XX
 PR 12-MAR-1999; 99US-0266965.
 XX
 PA (MINU) UNIV MINNESOTA.
 PA (SHER/) SHERMAN D H.
 PA (MAOY/) MAO Y.
 PA (VARO/) VAROGLU M.
 PA (HEM/) HE M.
 PA (SHEL/) SHELTON P C.
 XX
 PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
 XX
 DR WPI; 2000-601980/57.
 XX
 PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene
 PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
 PT the molecular basis of mitosome ring system biosynthesis -
 XX
 PS Claim 21; Figure 23; 399pp; English.
 XX
 CC This invention relates to isolated and purified nucleic acid molecules
 CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
 CC natural products that contain a variety of functional groups, including
 CC amino benzquinone and axiridine ring systems. The S. lavendulae
 CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
 CC spanning 55kb of DNA. The invention includes an expression cassette
 CC comprising a mitomycin biosynthetic gene operably linked to a promoter,
 CC and host cells transformed with the cassette. The nucleotide, and protein
 CC sequences and the transformed host cells of the invention result in
 CC antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
 CC antibiotic activities. The nucleotide sequences are used to elucidate the
 CC molecular basis for the biosynthesis of the mitosome ring system, as well
 CC as to engineer the biosynthesis of novel natural products, e.g.
 CC antibiotics, anti-inflammatory agents, anti-cancer agents,
 CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic
 CC obstructive pulmonary disease as well as other disease involving
 CC respiratory inflammation, or cholesterol-lowering agents or as crop
 CC protection agents (e.g. fungicides or insecticides) as well as
 CC biopolymers, e.g., in packaging or biomedical applications, or to engineer
 CC PHA monomer syntheses. Sequences AAC55782-C55881, AAC55815-C55849 and
 CC AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
 CC sequences and encoded proteins. Sequences AAC55812-C55814,
 CC AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
 CC cloning of the mitomycin biosynthetic genes.

SQ Sequence 18034 BP; 3203 A; 6345 C; 6150 G; 2335 T; 1 other;
 Query Match 45.0%; Score 23.4; DB 21; Length 18034;
 Best Local Similarity 73.2%; Pred. No. 2e+02;
 Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 9 CTCTCAGGCTCTCATCGCTCTCGTGGTCCGCGCTCGCCAGTG 49
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 11562 CGTGGCGGCTCATGGCCCTCGTGGCTTCGGCGCACCGCA 11602
 RESULT 37
 ABX30608/c
 ID ABX30608 standard; cDNA; 271 BP.
 XX
 AC ABX30608;
 XX
 DT 11-FEB-2003 (first entry)
 XX
 DE Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #12665.
 XX
 KW Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;
 KW cellular fucosylation; glycoconjugate fucosylation; transplant rejection;
 KW arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
 KW complex carbohydrate; gene replacement therapy; immunosuppressive;
 KW antiinflammatory; antiarthritic; antibacterial; cerebroprotective;
 KW antiasthmatic; vasotropic.
 XX
 OS Homo sapiens.
 XX
 PN US2002110548-A1.
 XX
 PD 15-AUG-2002.
 XX
 PF 11-JUN-2001; 2001US-0878574.
 XX
 PR 22-NOV-1996; 96US-0753233.
 PR 03-DEC-1997; 97US-0984246.
 PR 09-SEP-1998; 98US-0149674.
 PR 14-JUN-1999; 99US-0333177.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Sullivan F, Kriz R, Kumar R;
 XX
 DR WPI; 2003-066673/06.
 XX
 PS New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)
 PT peptide, for manufacturing complex carbohydrates, or as targets for
 PT screening GM4,6D antagonists for treating e.g. arthritis, or transplant
 PT rejection -
 XX
 CC Disclosure; SEQ ID NO 12667; 6pp; English.
 XX
 CC The invention relates to a composition comprising a human GDP-mannose
 CC 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
 CC GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
 CC in a mammalian subject and for treating or ameliorating diseases affected
 CC by the level of cellular fucosylation or diseases affected by the
 CC fucosylation of glycoconjugates. These diseases include arthritis,
 CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or
 CC infection. The GM4,6D peptide or a polynucleotide encoding it is also
 CC useful for manufacturing complex carbohydrates and as targets for
 CC screening small molecule antagonists of the activity of the enzyme. The
 CC polynucleotide is useful in developing an assay for defects in the
 CC enzyme, as well as in gene replacement therapy. Sequences
 CC ABX17942-ABX17944 and ABX17947-ABX17916 represent DNA molecules encoding
 CC human GM4,6D peptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 271 BP; 76 A; 49 C; 69 G; 76 T; 1 other;

Query Match 44.6%; Score 23.2; DB 25; Length 271;
Best Local Similarity 70.5%; Pred. No. 1.4e+02;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 8 TCTCTCAGCTCTCATCGCTCTCGTGGCGGCTCGCCAGT 51
DB 59 TCTGTCTGTCTCTCATCGCTCTCACTCCATTGTGGCCAGCCAGT 16

RESULT 38
ABQ98305/c
ID ABQ98305 standard; DNA; 450 BP.
XX
AC ABQ98305;
XX
DT 04-NOV-2002 (first entry)
XX
DE Human ORF112 coding sequence.
XX
KW Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;
KW Antinflammatory; gene therapy; human; ORFX; atherogenic; platelet;
KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;
KW cancer; cardiovascular disease; allergy; autoimmune disease;
KW wound healing; blood coagulation disorder; inflammatory disorder; ds.
XX
OS Homo sapiens.
XX
PN US2002082206-A1.
XX
PD 27-JUN-2002.
XX
PF 30-MAY-2001; 2001US-0867550.
XX
PR 30-MAY-2000; 2000US-208427P.
XX
XX (LEAC/) LEACH M D.
PA (MEHR/) MEHRABAN F.
PA (CONL/) CONLEY P B.
PA (TOPP/) TOPPER J N.
PA (LAWD/) LAW D.
XX
PI Leach MD, Mehraban F, Conley PB, Topper JN, Law D;
XX WPI; 2002-626554/67.
DR P-PSDB; ABP63742.
XX
PT New polypeptide designated ORFX are present in human atherogenic cells
PT and are useful to prevent and treat ORFX-associated disorders including
PT cancer, allergy, wound healing or autoimmune, cardiovascular or
PT inflammatory disease -
XX
PS Claim 2; SEQ ID 223; 78pp; English.
XX
XX The present invention relates to novel human ORFX polypeptides and their
XX coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences
XX were discovered in human atherogenic cells, in particular in platelets
XX and human umbilical vein endothelial cells (HUVEC) and are expressed in
XX many other tissues as well. Atherogenic cells are cells which have the
XX potential to develop atherosclerotic plaques. The ORFX polypeptides and
XX nucleic acids are useful for treating or preventing a pathological
XX condition associated with an ORFX-associated disorder, e.g. cancer,
XX cardiovascular disease, allergy, autoimmune disease, wound healing, blood
XX coagulation disorders or inflammatory disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USFTO web site at seqdata.uspto.gov/sequence.html?DocID=2002082206.
XX
XX Sequence 450 BP; 113 A; 113 C; 159 G; 62 T; 3 other;

Query Match 44.6%; Score 23.2; DB 24; Length 450;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 13 CACGCTCTCATCGCTCTCGTGGCGGCTCGCC 48
DB 207 CTCGATCTCGCGGCTCTCTCGCGGCTTCTCTCGCC 172

RESULT 39
ABL26959
ID ABL26959 standard; DNA; 867 BP.
XX
AC ABL26959;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 32350.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
PS Claim 1; SEQ ID NO 32350; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABBS7737-ABBS72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 867 BP; 220 A; 251 C; 207 G; 189 T; 0 other;

Query Match 44.6%; Score 23.2; DB 23; Length 867;
Best Local Similarity 65.4%; Pred. No. 1.6e+02;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ATGCAGTTTCTCTCACGCTCTCATCGCTCGCTCGCGGCTCGCCAGT 52
DB 636 AGGCAGTTTCTCTCGCGGCTCTCTCTCTCTCATGACACCTCGGCATTG 687

RESULT 40
ABL60268
ID ABL60268 standard; cDNA; 1563 BP.
XX
AC ABL60268;
XX
DT 23-AUG-2002 (first entry)
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 07:05:30 ; Search time 57 Seconds
(without alignments)
402.665 Million cell updates/sec

Title: US-10-081-935-1
Perfect score: 52
Sequence: 1 atgcagttctctcagctctt.....cgctgcggcctcgccagtg 52

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	48.1	651	4	US-09-252-991A-10364
2	25	48.1	2004	4	US-09-252-991A-10609
3	25	48.1	2691	4	US-09-252-991A-10245
4	23.8	45.8	49377	1	US-08-764-233A-1
5	23.4	45.0	972	4	US-09-252-991A-15702
6	23.4	45.0	975	4	US-09-252-991A-15674
7	23.4	45.0	1485	4	US-09-252-991A-15656
8	23.4	45.0	2466	4	US-09-266-965-83
9	23.4	45.0	18034	4	US-09-266-965-75
10	23.2	44.6	201	4	US-09-252-991A-782
11	23.2	44.6	1038	1	US-07-661-610C-13
12	23.2	44.6	1899	1	US-08-443-104-6
13	23.2	44.6	1899	1	US-08-442-859-6
14	23.2	44.6	1899	1	US-08-398-489-6
15	23.2	44.6	1899	5	PCT-US95-05534-6
16	23.2	44.6	2052	1	US-08-443-104-5
17	23.2	44.6	2052	1	US-08-238-130-6
18	23.2	44.6	2052	1	US-08-442-859-5
19	23.2	44.6	2052	2	US-08-398-489-5
20	23.2	44.6	2052	2	US-08-894-772-1
21	23.2	44.6	2052	2	US-09-207-844-1
22	23.2	44.6	2052	4	PCT-US95-05534-5
23	23.2	44.6	2052	5	PCT-US95-05534-5
24	23.2	44.6	5437	1	US-07-661-610C-1
25	23.2	44.6	5762	4	US-09-647-492A-2
26	23	44.2	3903	6	520236-12
27	23	44.2	319608	4	US-09-539-333D-1

c	28	23	44.2	319608	4	US-09-679-409-1	Sequence 1, Appli
	29	22.4	43.1	48	2	US-08-169-948B-35	Sequence 35, Appli
	30	22.4	43.1	48	2	US-08-448-873-35	Sequence 35, Appli
	31	22.4	43.1	48	3	US-08-382-452D-35	Sequence 35, Appli
	32	22.4	43.1	49	4	US-08-507-362A-19	Sequence 19, Appli
	33	22.4	43.1	702	3	US-09-146-770-2	Sequence 2, Appli
	34	22.4	43.1	702	3	US-09-216-295-2	Sequence 2, Appli
	35	22.4	43.1	702	4	US-09-633-084-2	Sequence 2, Appli
	36	22.4	43.1	702	4	US-10-075-872-2	Sequence 2, Appli
	37	22.4	43.1	702	4	US-10-261-997-2	Sequence 2, Appli
	38	22.4	43.1	1050	1	US-08-032-848C-11	Sequence 11, Appli
	39	22.4	43.1	1050	1	US-08-438-870-11	Sequence 11, Appli
	40	22.4	43.1	1050	2	US-08-774-065-1	Sequence 1, Appli
c	41	22.4	43.1	1365	4	US-09-252-991A-8265	Sequence 8265, Ap
	42	22.2	42.7	703	3	US-08-998-416-751	Sequence 751, App
	43	22.2	42.7	1140	1	US-08-289-653-2	Sequence 2, Appli
	44	22.2	42.7	2128	4	US-09-620-312D-197	Sequence 197, App
	45	22.2	42.7	2845	1	US-08-289-653-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-252-991A-10364
; Sequence 10364, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10364
; LENGTH: 651
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10364

Query Match 48.1%; Score 25; DB 4; Length 651;
Best Local Similarity 69.4%; Pred. No. 7.1;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 4 CAGTTCTCTCAGCTCTCATCGCTCTCGCTCGCGCGCTCGCCAGTG 52
Db 300 CAGTCGCTGACCTCTCTCTGCTCGCTCGCGCGCGCGCTCGCGC 348

RESULT 2
US-09-252-991A-10609/c
; Sequence 10609, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10609
; LENGTH: 2004
; TYPE: DNA

```
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10609

Query Match      48.1%; Score 25; DB 4; Length 2004;
Best Local Similarity 69.4%; Pred. No. 8.1;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY      4 CAGTTCTCTCAGCGCTCTCATCGCTGCTGCTGCCGCCGCTCGCAGT 52
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      591 CAGCTCGCTGACCTCCTCTTGGCTTGTGCGACGCCGCCGACGTCACGC 543

RESULT 3
US-09-252-991A-10245
; Sequence 10245, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10245
; LENGTH: 2691
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10245

Query Match      48.1%; Score 25; DB 4; Length 2691;
Best Local Similarity 69.4%; Pred. No. 8.4;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY      4 CAGTTCTCTCAGCGCTCTCATCGCTGCTGCTGCCGCCGCTCGCAGT 52
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2161 CAGCTCGCTGACCTCCTCTTGGCTTGTGCGACGCCGCCGACGTCACGC 2209

RESULT 4
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:
; CLONE: p98/1, pUL3, and pVKM15
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 383..760
; OTHER INFORMATION: /product= "Sorr"
; OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologo
; OTHER INFORMATION: the reductase domains of type I PKSs such as eryA from
; OTHER INFORMATION: Saccharopolyspora erythraea."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 927..19874
; OTHER INFORMATION: /product= "Sora"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSs th
; OTHER INFORMATION: are known to be involved in the synthesis of polyketide
; OTHER INFORMATION: compounds."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 942..7115
; OTHER INFORMATION: /product= "Module 1 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7203..12884
; OTHER INFORMATION: /product= "Module 2 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13455..19616
; OTHER INFORMATION: /product= "Module 3 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19870..24556
; OTHER INFORMATION: /product= "Module 1 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 24638..30820
; OTHER INFORMATION: /product= "Module 2 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30881..35446
; OTHER INFORMATION: /product= "Module 3 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 35528..40114
; OTHER INFORMATION: /product= "Module 4 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 40190..46318
```

```
/ OTHER INFORMATION: /product= "Module 5 of SorB"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 46851..47891
/ OTHER INFORMATION: /product= "SorM"
/ OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly
/ OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
/ OTHER INFORMATION: hyposcopus that is involved in the synthesis of the
/ OTHER INFORMATION: polyketide rappamycin."
US-08-764-233A-1

Query Match          45.8%; Score 23.8; DB 1; Length 49377;
Best Local Similarity 72.1%; Pred. No. 29;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 6 GTTCTCAGCGCTCATGCTCTCGTGGCTGGCGCTGCC 48
Db 5048 GTTCACCGCCGCGCTCGCGCTCTCCAGCTCGAGCGGCTCACC 5090

RESULT 5
US-09-252-991A-15702
/ Sequence 15702; Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252.991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 15702
/ LENGTH: 972
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (669)
/ OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-15702

Query Match          45.0%; Score 23.4; DB 4; Length 972;
Best Local Similarity 73.2%; Pred. No. 25;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ATGCACTTCTCTCAGCGCTCATGCTCTCGTGGCTGCCG 41
Db 151 ATGCAGGCGCGCCGCCCTGATCGCGCTGTCGGCGCCG 191

RESULT 6
US-09-252-991A-15674
/ Sequence 15674; Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252.991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 15674
/ LENGTH: 975

TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: unsure
LOCATION: (571)
OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-15674

Query Match          45.0%; Score 23.4; DB 4; Length 975;
Best Local Similarity 73.2%; Pred. No. 25;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ATGCACTTCTCTCAGCGCTCATGCTCTCGTGGCTGCCG 41
Db 53 ATGCAGGCGCGCCGCCCTGATCGCGCTGTCGGCGCCG 93

RESULT 7
US-09-252-991A-15656/c
/ Sequence 15656; Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252.991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 15656
/ LENGTH: 1485
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (415)
/ OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-15656

Query Match          45.0%; Score 23.4; DB 4; Length 1485;
Best Local Similarity 73.2%; Pred. No. 26;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ATGCACTTCTCTCAGCGCTCATGCTCTCGTGGCTGCCG 41
Db 933 ATGCAGGCGCGCCGCCCTGATCGCGCTGTCGGCGCCG 893

RESULT 8
US-09-266-965-83
/ Sequence 83; Application US/09266965
/ Patent No. 6495348
/ GENERAL INFORMATION:
/ APPLICANT: Sherman, D
/ APPLICANT: Mao, Y
/ APPLICANT: Varoglu, M
/ APPLICANT: He, M
/ APPLICANT: Sheldon, P
/ TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
/ FILE REFERENCE: 600.456US1
/ CURRENT APPLICATION NUMBER: US/09/266.965
/ PRIOR FILING DATE: 1999-03-12
/ PRIOR APPLICATION NUMBER: US 08/624,447
/ PRIOR FILING DATE: 1996-08-19
/ PRIOR APPLICATION NUMBER: PCT/US94/11279
/ PRIOR FILING DATE: 1994-10-06
/ PRIOR APPLICATION NUMBER: US 08/133,963
/ PRIOR FILING DATE: 1993-10-07
/ NUMBER OF SEQ ID NOS: 145
```

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 83
LENGTH: 2466
TYPE: DNA
ORGANISM: Streptomyces lavendulae
US-09-266-965-83

Query Match 45.0%; Score 23.4; DB 4; Length 2466;
Best Local Similarity 73.2%; Pred. No. 28;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 9 CTCACGCTCTCATCGCTCTCGTCCGCCCTGCCA 49
| | | | | | | | | | | | | | | | | | | | | |
Db 48 CGCTGCCGCCCTGATGCGCTCTCGTCCGCCGCCA 88

RESULT 9
US-09-266-965-75
Sequence 75, Application US/09266965
Patent No. 6495348
GENERAL INFORMATION:
APPLICANT: Sherman, D
APPLICANT: Mao, Y
APPLICANT: Varoglu, M
APPLICANT: He, M
APPLICANT: Sheldon, P
TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
FILE REFERENCE: 600,456US1
CURRENT APPLICATION NUMBER: US/09/266,965
CURRENT FILING DATE: 1999-03-12
EARLIER APPLICATION NUMBER: US 08/624,447
EARLIER FILING DATE: 1996-08-19
EARLIER APPLICATION NUMBER: PCT/US94/11279
EARLIER FILING DATE: 1994-10-06
EARLIER APPLICATION NUMBER: US 08/33,963
EARLIER FILING DATE: 1993-10-07
NUMBER OF SEQ ID NOS: 145
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 75
LENGTH: 18034
TYPE: DNA
ORGANISM: Streptomyces lavendulae
FEATURE:
NAME/KEY: unsure
LOCATION: (302)... (302)
OTHER INFORMATION: n is a or t or g or c
US-09-266-965-75

Query Match 45.0%; Score 23.4; DB 4; Length 18034;
Best Local Similarity 73.2%; Pred. No. 35;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 9 CTCACGCTCTCATCGCTCTCGTCCGCCCTGCCA 49
| | | | | | | | | | | | | | | | | | | | | |
Db 11562 CGCTGCCGCCCTGATGCGCTCTCGTCCGCCGCCA 11602

RESULT 10
US-09-252-991A-782/C
Sequence 782, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 31142

SEQ ID NO 782
LENGTH: 201
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-782

Query Match 44.6%; Score 23.2; DB 4; Length 201;
Best Local Similarity 70.5%; Pred. No. 24;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 CAGTTCTCAGCGCTCATCGCTCTCGTCCGCCCTGCC 47
| | | | | | | | | | | | | | | | | | | | | |
Db 135 CTGTCGCTCATGCTCTGATGCTTCTGCACATGCCGTCGC 92

RESULT 11
US-07-661-610C-13
Sequence 13, Application US/07661610C
Patent No. 5292643
GENERAL INFORMATION:
APPLICANT: Shibano, Yuji
APPLICANT: Toyoda, Hideyoshi
APPLICANT: Utsumi, Ryutaro
APPLICANT: Obata, Kazuaki
TITLE OF INVENTION: Fusaric Acid Resistant Genes
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/661,610C
FILING DATE: 19910228
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5292643man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2292-010-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-07-661-610C-13

Query Match 44.6%; Score 23.2; DB 1; Length 1038;
Best Local Similarity 65.4%; Pred. No. 29;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ATGCAAGTTCTCAGCGCTCATCGCTCTCGTCCGCCCTGCCAGTG 52
| | | | | | | | | | | | | | | | | | | | | |
Db 13 ACGGAGGTCTCCAGATCATCATCGGATGCTGTCGCCGCCGCTGCACGC 64

RESULT 12
US-08-443-104-6
Sequence 6, Application US/08443104
Patent No. 5691162
GENERAL INFORMATION:

APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 56911620 No. 5691162disk of No. 5691162th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443,104
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1899 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-443-104-6

Query Match 44.6%; Score 23.2; DB 1; Length 1899;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1 ATGCGTTCTCTCAGCGCTCTCATCGCTCTCGTGGCTGCCGCGCTGCCAGTG 52
Db 1 ATGCGTTCTCTCAGCGCTCTCTCTCTCATCGGCTATCCAGCGCTCGCTGGTG 52

RESULT 13
US-08-442-859-6
Sequence 6, Application US/08442859
Patent No. 5807729
GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58077290 No. 5807729disk of No. 5807729th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:

MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/442,859
FILING DATE: 17-MAY-1995
CLASSIFICATION: 252
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1899 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-442-859-6

Query Match 44.6%; Score 23.2; DB 1; Length 1899;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1 ATGCGTTCTCTCAGCGCTCTCATCGCTCTCGTGGCTGCCGCGCTGCCAGTG 52
Db 1 ATGCGTTCTCTCAGCGCTCTCTCTCTCATCGGCTATCCAGCGCTCGCTGGTG 52

RESULT 14
US-08-398-489-6
Sequence 6, Application US/08398489
Patent No. 5843753
GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58437530 No. 5843753disk of No. 5843753th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/398,489
FILING DATE: 03-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 1899 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-398-489-6

Query Match 44.6%; Score 23.2; DB 2; Length 1899;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ATGCACTTCTCAGCCTCTCATGCTCTGTCGTCGCCGCCCTGCCAGTG 52
DB 1 ATGCGTTCTCCGACTCTCTCCTCATCGGCTATCCAGCTCGCTG 52

RESULT 15
PCT-US95-05534-6
Sequence 6, Application PC/TUS9505534
GENERAL INFORMATION:

APPLICANT:
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Novo Nordisk of North America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05534
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238/108
FILING DATE: 04-MAY-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489
FILING DATE: 03-MARCH-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Agtis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.204-MO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1899 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US95-05534-6

Query Match 44.6%; Score 23.2; DB 5; Length 1899;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ATGCACTTCTCAGCCTCTCATGCTCTGTCGTCGCCGCCCTGCCAGTG 52
DB 1 ATGCGTTCTCCGACTCTCTCCTCATCGGCTATCCAGCTCGCTG 52

RESULT 16
US-08-443-104-5
Sequence 5, Application US/08443104
Patent No. 5691162
GENERAL INFORMATION:

APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 56911620 No. 5691162disk of No. 5691162th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443,104
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agtis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-443-104-5

Query Match 44.6%; Score 23.2; DB 1; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ATGCACTTCTCAGCCTCTCATGCTCTGTCGTCGCCGCCCTGCCAGTG 52
DB 1 ATGCGTTCTCCGACTCTCTCCTCATCGGCTATCCAGCTCGCTG 52

RESULT 17
US-08-238-130-6
Sequence 6, Application US/08238130
Patent No. 5702934
GENERAL INFORMATION:
APPLICANT: Hastud, Sven
APPLICANT: Branner, Sven
APPLICANT: Jorgensen, Birthe R.
APPLICANT: Christensen, Tove
APPLICANT: Jorgensen, Birgitte B.
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus

TITLE OF INVENTION: PROCESSES FOR PRODUCING AN ENZYME
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 57029340 No. 5702934disk of No. 5702934th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/238,130
FILING DATE: 04-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 522/93
FILING DATE: 05-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr. Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 3965.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(1..363, 416..802, 856..1821, 1870..2052)
FEATURE:
NAME/KEY: mat peptide
LOCATION: join(1..363, 416..802, 856..1821, 1870..2049)
US-08-238-130-6
Query Match 44.6%; Score 23.2; DB 1; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1 ATGCAGTTCTCCAGCCTCTCATCGCTCGTCGCGCCGCTCCCACTG 52
DB 1 ATGCCTTCTCCGACTCTCTCCTCATCGGCTATCCAGCTCGCTGTG 52
RESULT 18
US-08-442-859-5
Sequence 5, Application US/08442859
GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58077290 No. 5807729disk of No. 5807729th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:

MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/442,859
FILING DATE: 17-MAY-1995
CLASSIFICATION: 252
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-442-859-5
Query Match 44.6%; Score 23.2; DB 1; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1 ATGCAGTTCTCCAGCCTCTCATCGCTCGTCGCGCCGCTCCCACTG 52
DB 1 ATGCCTTCTCCGACTCTCTCCTCATCGGCTATCCAGCTCGCTGTG 52
RESULT 19
US-08-398-489-5
Sequence 5, Application US/08398489
GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58437530 No. 5843753disk of No. 5843753th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/398,489
FILING DATE: 03-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-398-489-5

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

RESULT 20

US-08-894-772-1
Sequence 1, Application US/08894772
Patent No. 5861280

GENERAL INFORMATION:
APPLICANT: Lehmebeck, Jan

TITLE OF INVENTION: Host Cell Expressing Reduced Levels
TITLE OF INVENTION: Of A Metalloprotease And Methods Using The Host Cell In Protei

TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58612800 No. 5861280disk of No. 5861280th America, Inc.

STREET: 405 Lexington Avenue, Suite 6400
CITY: New York

STATE: New York
COUNTRY: U.S.A.

ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894.772

FILING DATE: 27-AUG-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Gregg, Valetta A.
REGISTRATION NUMBER: 35,127

REFERENCE/DOCKET NUMBER: 4300.204-US
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 2052 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-894-772-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-894-772-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-894-772-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-894-772-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-894-772-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Sequence 1, Application US/09207844
Patent No. 5968774
GENERAL INFORMATION:
APPLICANT: Lehmebeck, Jan
TITLE OF INVENTION: Host Cell Expressing Reduced Levels
TITLE OF INVENTION: Of A Metalloprotease And Methods Using The Host Cell In Protei

TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59687740 No. 5968774disk of No. 5968774th America, Inc.

STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.

ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/207,844

FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/894,772

FILING DATE: 27-AUG-1997
ATTORNEY/AGENT INFORMATION:

NAME: Gregg, Valetta A.
REGISTRATION NUMBER: 35,127

REFERENCE/DOCKET NUMBER: 4300.204-US
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 2052 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-09-207-844-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-09-207-844-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-09-252-509-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-09-252-509-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-09-252-509-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-09-252-509-1

Query Match 44.6%; Score 23.2; DB 2; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 1 ATGCAGTTCTCTCAGCTTCATGCTCTGCTCGCTCGCGGCTCGGCAGTG 52
Db 1 ATGCAGTTCTCTCAGCTCTCTCTCTCTCATCGGCTATCCAGCTCGCTGGTG 52

FEATURE:
NAME/KEY: mat peptide
LOCATION: (785)...(2049)
NAME/KEY: sig_peptide
LOCATION: (55)...(784)
NAME/KEY: intcon
LOCATION: (364)...(415)
NAME/KEY: intcon
LOCATION: (802)...(854)
NAME/KEY: intcon
LOCATION: (1821)...(1868)
US-09-252-509-1

Query Match 44.6%; Score 23.2; DB 4; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1 ATGCAGTTCTTCACGCTCTCATGCTCTGTCGTCGCGGCTCCCACTG 52
Db 1 ATGCGTTCTCCGACTCTCTCCCTCATCGGCTATCCAGCTCCCTGCTG 52

RESULT 23

PCT-US95-05534-5
Sequence 5, Application PC/TUS9505534
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: NOVEL METALLOPROTEIN HAVING INCREASED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Novo Nordisk of North America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05534
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238/108
FILING DATE: 04-MAY-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489
FILING DATE: 03-MARCH-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr. Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.204-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US95-05534-5

Query Match 44.6%; Score 23.2; DB 5; Length 2052;
Best Local Similarity 65.4%; Pred. No. 32;

Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1 ATGCAGTTCTTCACGCTCTCATGCTCTGTCGTCGCGGCTCCCACTG 52
Db 1 ATGCGTTCTCCGACTCTCTCCCTCATCGGCTATCCAGCTCCCTGCTG 52

RESULT 24

US-07-661-610C-1
Sequence 1, Application US/07661610C
Patent No. 5292643
GENERAL INFORMATION:
APPLICANT: Shibano, Yuji
APPLICANT: Toyoda, Hideyoshi
APPLICANT: Utsunomiya, Ryutaro
APPLICANT: Obata, Kazuaki
TITLE OF INVENTION: Fused Acid Resistant Genes
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/661,610C
FILING DATE: 19910228
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5292643man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2292-010-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5437 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Pseudomonas cepacia
FEATURE:
NAME/KEY: RBS
LOCATION: 519..522
FEATURE:
NAME/KEY: CDS
LOCATION: 531..1832
FEATURE:
NAME/KEY: RBS
LOCATION: 1836..1839
FEATURE:
NAME/KEY: CDS
LOCATION: 1847..2275
FEATURE:
NAME/KEY: RBS
LOCATION: 2301..2304
FEATURE:
NAME/KEY: CDS
LOCATION: 2314..3354
FEATURE:
NAME/KEY: RBS
LOCATION: 3405..3409
FEATURE:

:	NAME/KEY:	CDS
:	LOCATION:	3414..4040
:	FEATURE:	
:	NAME/KEY:	RBS
:	LOCATION:	4457..4462
:	FEATURE:	
:	NAME/KEY:	CDS
:	LOCATION:	4473..4979
:		
US-07-661-610C-1		

Query Match	44.6%	Score 23.2;	DB 1;	Length 5437;
Best Local Similarity	65.4%;	Pred. No. 36;		
Matches 34;	Conservative 0;	Mismatches 18;	Indels 0;	Gaps 0;

QY 1 ATGAGTTCCTCACGCTCATCGCTCTCGTCGGCCGCCATG 52
 | | | | | | | | | | | | | | | |
Db 2326 ACGCGGTCTCGAAGTCATCATCGGATCGTGCGCCGGCTGTCAGCG 2377

```

RESULT 25
US-09-647-492A-2
; Sequence 2, Application US/09647492A
; Patent No. 6551797
; GENERAL INFORMATION:
; APPLICANT: PFALLER, Rupert
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR PRODUCTION OF PROTEINS
; FILE REFERENCE: PFALLER ET AL.-2 (PC)
; CURRENT APPLICATION NUMBER: US/09/647,492A
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 5762
; TYPE: DNA
; ORGANISM: Trameetes versicolor
US-09-647-492A-2

```

Query Match	44.6%	Score 23.2;	DB 4;	Length 5762;
Best Local Similarity	65.4%;	Pred. No. 36;		
Matches 34;	Conservative	0;	Mismatches 18;	Indels 0;
			Gaps	0;

Oy 1 ATGCAGTTCTCTACGCTGTCAATCGCTCTGTCGGTCCGGGCTCGCACAATG 52
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 ATGTCCAGGTTTTCACACTCTTCTTCGGCTTTCGTGCTTCCCTTGCGGCTG 599

RESULT 26
5202236-12/C
Patent No. 5202236
APPLICANT: MARG, KATHY J.; ANDERSON, DAVID M.; STRAUSBERG,
SUSAN L.; MCCANDLISH, RUSSETT, TENA; FILIPULA, DAVID
TITLE OF INVENTION: METHOD OF PRODUCING BIOADHESIVE
PROTEIN
NUMBER OF SEQUENCES: 3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/528,762
FILING DATE: 25-MAY-1990
APPLICATION NUMBER: 82,456
FILING DATE: 07-AUG-1987
APPLICATION NUMBER: 933,945
FILING DATE: 24-NOV-1986
APPLICATION NUMBER: 650,128
FILING DATE: 13-SEP-1984
SEQ ID NO:12
LENGTH: 3903
5202236-12

```
Query Match      44.2%; Score 23; DB 6; Length 3903;  
Best Local Similarity    74.4%;   Pred. No. 40;  
Matches          29; Conservative    0; Mismatches 10; Indels    0;  
  
QY       11 CTCACGCTTCATCGCCTCTCCGCCGCAGCGCCGCTGGCCA 49  
| | ||||| ||||| | | | ||||| |||||
```

11 CTACGCTCATCGCTCTCGTGGCTGCCGCCCTGCCA 49

Db 1025 CGCCCGCTCTTCGCTTCTTCCCTTCCGGCTCGCA 987

```

RESULT 27
US-09-539-333D-1/c
Sequence 1: Application US/09539333D
Patent No. 6476208
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Matra
APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Lydie
APPLICANT: Bihain, Bernard
APPLICANT: Essieux, Laurent
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLERIC MARKERS
FILE REFERENCE: GENSET.047AUS
CURRENT FILING DATE: 2000-03-10
PRIORITY APPLICATION NUMBER: US 60/126,903
PRIORITY FILING DATE: 1999-03-30
PRIORITY APPLICATION NUMBER: US 60/131,971
PRIORITY FILING DATE: 1999-04-30
PRIORITY APPLICATION NUMBER: US 60/132,065
PRIORITY FILING DATE: 1999-04-30
PRIORITY APPLICATION NUMBER: US 60/143,928
PRIORITY FILING DATE: 1999-07-14
PRIORITY APPLICATION NUMBER: US 60/145,915
PRIORITY FILING DATE: 1999-07-27
PRIORITY APPLICATION NUMBER: US 60/146,453
PRIORITY FILING DATE: 1999-07-29
PRIORITY APPLICATION NUMBER: US 60/146,452
PRIORITY FILING DATE: 1999-07-29
PRIORITY APPLICATION NUMBER: US 60/162,288
PRIORITY FILING DATE: 1999-10-28
PRIORITY APPLICATION NUMBER: US 09/416,384
PRIORITY FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 231
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 319608
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 31..1107
OTHER INFORMATION: 5' regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 1108..1289
OTHER INFORMATION: exon A g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 14877..14920
OTHER INFORMATION: exon B g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 18778..18862
OTHER INFORMATION: exon Bbis g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 25593..25740
OTHER INFORMATION: exon C g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29388..29502
OTHER INFORMATION: exon D g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29967..30282
OTHER INFORMATION: exon E g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 64666..64812

```

OTHER INFORMATION: exon F g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 65505..65853
OTHER INFORMATION: exon G g35018 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 65854..67854
OTHER INFORMATION: 3'regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 94124..94964
OTHER INFORMATION: exon g35017
FEATURE:
NAME/KEY: exon
LOCATION: 201188..201234
OTHER INFORMATION: exon S g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 214676..214793
OTHER INFORMATION: exon T g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215702..215746
OTHER INFORMATION: exon U g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216836..216915
OTHER INFORMATION: exon V g35030 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 213818..215818
OTHER INFORMATION: 3'regulatory region g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215819..215941
OTHER INFORMATION: exon R complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215819..215975
OTHER INFORMATION: exon Rbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216661..216952
OTHER INFORMATION: exon Qbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216661..217061
OTHER INFORMATION: exon Q complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 217027..217061
OTHER INFORMATION: exon Q1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 229647..229742
OTHER INFORMATION: exon X complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 230408..230721
OTHER INFORMATION: exon P complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231272..231412
OTHER INFORMATION: exon Obis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231787..231880
OTHER INFORMATION: exon O2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231870..231879
OTHER INFORMATION: exon O1 complement g34872 gene

FEATURE:
NAME/KEY: exon
LOCATION: 234174..234321
OTHER INFORMATION: exon O complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 237406..237428
OTHER INFORMATION: exon Nbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239807
OTHER INFORMATION: exon N2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239853
OTHER INFORMATION: exon N complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240569
OTHER INFORMATION: exon M1117 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240596
OTHER INFORMATION: exon M1090 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240617
OTHER INFORMATION: exon M1069 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240644
OTHER INFORMATION: exon M52 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240824
OTHER INFORMATION: exon M862 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..241685
OTHER INFORMATION: exon M1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240800..240993
OTHER INFORMATION: exon M51 complement g34872 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 241686..243685
OTHER INFORMATION: 5'regulatory region g34872 gene
FEATURE:
NAME/KEY: misc feature
LOCATION: 290652..292652
OTHER INFORMATION: 3'regulatory region g34665 gene
FEATURE:
NAME/KEY: exon
LOCATION: 292653..292841

Query Match 44.2% Score 23; DB 4; Length 319608;
Best local Similarity 68.1%; Pred. No. 65;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 4 CAGTTCCTCAGCGCTCATCGCTCTGTCGTCGCGCGCTCCGCGAG 50
Db 184156 CAGTTCACTAAGGACACTCAGCTCTGCTTGGCTGTGATACCCAG 184110

RESULT 28
US-09-679-409-1/c
; Sequence 1, Application US/09679409
; Patent No. 6555316

```
/ GENERAL INFORMATION:
/ APPLICANT: Cohen, Daniel
/ APPLICANT: Blumenfeld, Marla
/ APPLICANT: Chumakov, Ilya
/ APPLICANT: Bougueleret, Lydie
/ APPLICANT: Essioux, Laurent
/ TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
/ FILE REFERENCE: 53.US15.CIP
/ CURRENT APPLICATION NUMBER: US/09/679,409
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 09/539,333
/ PRIOR FILING DATE: 2000-03-03
/ PRIOR APPLICATION NUMBER: 09/416,384
/ PRIOR FILING DATE: 1999-10-12
/ PRIOR APPLICATION NUMBER: 60/168,088
/ PRIOR FILING DATE: 1999-11-30
/ NUMBER OF SEQ. ID NOS: 134
/ SOFTWARE: Patent.pm
/ SEQ ID NO 1
/ LENGTH: 3,9608
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: 199122..201122
/ OTHER INFORMATION: 5'regulatory region
/ NAME/KEY: exon
/ LOCATION: 201123..201234
/ OTHER INFORMATION: exon S
/ NAME/KEY: exon
/ LOCATION: 201123..201560
/ OTHER INFORMATION: exon S2
/ NAME/KEY: exon
/ LOCATION: 214676..214793
/ OTHER INFORMATION: exon T
/ NAME/KEY: exon
/ LOCATION: 215702..215746
/ OTHER INFORMATION: exon U
/ NAME/KEY: exon
/ LOCATION: 216836..216994
/ OTHER INFORMATION: exon V
/ NAME/KEY: exon
/ LOCATION: 216836..217077
/ OTHER INFORMATION: exon V2
/ NAME/KEY: exon
/ LOCATION: 217671..217764
/ OTHER INFORMATION: exon V1
/ NAME/KEY: exon
/ LOCATION: 227655..227736
/ OTHER INFORMATION: exon V4
/ NAME/KEY: exon
/ LOCATION: 238715..238919
/ OTHER INFORMATION: exon V3
/ NAME/KEY: exon
/ LOCATION: 240440..240673
/ OTHER INFORMATION: exon W
/ NAME/KEY: exon
/ LOCATION: 240440..241153
/ OTHER INFORMATION: exon W2
/ NAME/KEY: exon
/ LOCATION: 241072..241291
/ OTHER INFORMATION: exon X
/ NAME/KEY: exon
/ LOCATION: 244353..244561
/ OTHER INFORMATION: exon Y
/ NAME/KEY: exon
/ LOCATION: 246273..247802
/ OTHER INFORMATION: exon Z
/ NAME/KEY: misc_feature
/ LOCATION: 247803..249803
/ OTHER INFORMATION: 3'regulatory region
/ NAME/KEY: allele
/ LOCATION: 8316

/ OTHER INFORMATION: 99-27943-150 : polymorphic base G or C
/ NAME/KEY: allele
/ LOCATION: 21672
/ OTHER INFORMATION: 99-27935-193 : polymorphic base G or C
/ NAME/KEY: allele
/ LOCATION: 65485
/ OTHER INFORMATION: 8-128-33 : polymorphic base C or T
/ NAME/KEY: allele
/ LOCATION: 95396
/ OTHER INFORMATION: 99-31960-363 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 107281
/ OTHER INFORMATION: 99-24656-260 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 160640
/ OTHER INFORMATION: 99-24639-163 : polymorphic base A or C
/ NAME/KEY: allele
/ LOCATION: 160876
/ OTHER INFORMATION: 99-24634-108 : polymorphic base A or T
/ NAME/KEY: allele
/ LOCATION: 168974
/ OTHER INFORMATION: 99-7652-162 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 170810
/ OTHER INFORMATION: 99-16100-147 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 173358
/ OTHER INFORMATION: 99-5862-167 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 189957
/ OTHER INFORMATION: 99-5919-215 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 197163
/ OTHER INFORMATION: 99-24658-410 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 200778
/ OTHER INFORMATION: 8-303-235 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 202651
/ OTHER INFORMATION: 8-300-221 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 202679
/ OTHER INFORMATION: 8-300-193 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 203378
/ OTHER INFORMATION: 8-299-128 : polymorphic base A or T
/ NAME/KEY: allele
/ LOCATION: 204138
/ OTHER INFORMATION: 8-296-213 : polymorphic base A or T
/ NAME/KEY: allele
/ LOCATION: 204605
/ OTHER INFORMATION: 8-252-190 : polymorphic base C or T
/ NAME/KEY: allele
/ LOCATION: 204934
/ OTHER INFORMATION: 99-24644-194 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 205206
/ OTHER INFORMATION: 8-295-248 : polymorphic base A or C
/ NAME/KEY: allele
/ LOCATION: 205329
/ OTHER INFORMATION: 8-295-125 : polymorphic base C or T
/ NAME/KEY: allele
/ LOCATION: 206064
/ OTHER INFORMATION: 8-293-130 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 206545
/ OTHER INFORMATION: 8-292-198 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 207313
/ OTHER INFORMATION: 8-251-322 : polymorphic base A or G
/ NAME/KEY: allele
/ LOCATION: 208285
/ OTHER INFORMATION: 8-289-322 : polymorphic base A or G
```

```

; NAME/KEY: allele
; LOCATION: 208960
; OTHER INFORMATION: 8-287-249 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 209123
; OTHER INFORMATION: 8-287-86 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 209631
; OTHER INFORMATION: 8-285-319 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 210361
; OTHER INFORMATION: 8-283-278 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 210463
; OTHER INFORMATION: 8-283-176 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 210486
; OTHER INFORMATION: 8-283-153 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 210583
; OTHER INFORMATION: 8-283-56 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 210879
; OTHER INFORMATION: 8-282-345 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 210964
; OTHER INFORMATION: 8-282-260 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 210979
; OTHER INFORMATION: 8-282-245 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 211050
; OTHER INFORMATION: 8-282-174 : variable motif AAAG or GAAGGAAGGAAGGAAGA
; NAME/KEY: allele
; LOCATION: 211132
; OTHER INFORMATION: 8-282-92 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 211247
; OTHER INFORMATION: 8-281-367 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 211315
; OTHER INFORMATION: 8-281-299 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 211366
; OTHER INFORMATION: 8-281-248 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 212520
; OTHER INFORMATION: 8-279-197 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 212821
; OTHER INFORMATION: 8-278-289 : polymorphic base C or T
; NAME/KEY: allele

Query Match 44.2%; Score 23; DB 4; Length 319608;
Best Local Similarity 68.1%; Pred. No. 65;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 4 CAGTCTCTCAGCGCTCATCGCTCTCGCGCGCGCGCTCGCCAG 50
Db 184156 CAGTTCCTAAGGCACTCAGCTCTGCTGCTGCTATACCCAG 184110

RESULT 29
US-08-169-9488-35
; Sequence 35, Application US/081699488
; Patent No. 5861271
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; APPLICANT: Ward, Michael
; APPLICANT: Clarkson, Kathleen
; APPLICANT: Collier, Katherine
; APPLICANT: Larenas, Edmund
; TITLE OF INVENTION: No. 5861271e1 Cellulase Enzymes and Systems
```

```

; TITLE OF INVENTION: For Their Expression
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,948B
; FILING DATE: DEC 17 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-169-9488-35

Query Match 43.1%; Score 22.4; DB 2; Length 48;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Oy 1 ATGCAAGTTCTTCACGCTCTCATCGCTCTGCTGCGCGCGCTCGCC 48
Db 1 ATGCAAGTTCTTCACGCTCTCTCGCTCATACCGCGCGCGCTCGCC 48

RESULT 30
US-08-448-873-35
; Sequence 35, Application US/08448873
; Patent No. 5874276
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; APPLICANT: Ward, Michael
; APPLICANT: Clarkson, Kathleen
; APPLICANT: Collier, Katherine A.
; APPLICANT: Larenas, Edmund
; TITLE OF INVENTION: No. 5874276e1 Cellulase Enzymes and Systems
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,873
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: 08/169,948
FILING DATE: 17-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stone, Christopher L.
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC226D14
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 742-7555
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-448-873-35

Query Match      43.1%; Score 22.4; DB 2; Length 48;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY      1 ATGCAGTTCTCTCAGCCTCTCATGCTCTGCTGCGGCGGCTTGCC 48
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1 ATGAAGTTCCTTCAAGTCTCTCCCTGCTCATACCGGCGGCTTGCC 48

RESULT 31
US-08-382-452D-35
Sequence 35, Application US/08382452D
Patent No. 6268196
GENERAL INFORMATION:
APPLICANT: Fowler, Timothy
APPLICANT: Clarkson, Kathleen A.
APPLICANT: Ward, Michael
APPLICANT: Collier, Katherine D.
APPLICANT: Larenas, Edmund A.
TITLE OF INVENTION: NOVEL CELLULOSE ENZYMES AND SYSTEMS
TITLE OF INVENTION: FOR THEIR EXPRESSION
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESS: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/382,452D
FILING DATE: February 1, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 36,696
REFERENCE/DOCKET NUMBER: GC226-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 742-7555
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-382-452D-35

Query Match      43.1%; Score 22.4; DB 3; Length 48;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY      1 ATGCAGTTCTCTCAGCCTCTCATGCTCTGCTGCGGCGGCTTGCC 48
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1 ATGAAGTTCCTTCAAGTCTCTCCCTGCTCATACCGGCGGCTTGCC 48

RESULT 32
US-08-507-362A-19
Sequence 19, Application US/08507362A
Patent No. 6562340
GENERAL INFORMATION:
APPLICANT: Bedford, Michael
APPLICANT: Morgan, Andrew
APPLICANT: Fowler, Timothy
APPLICANT: Ward, Michael
APPLICANT: Clarkson, Kathleen
APPLICANT: Collier, Katherine
APPLICANT: Larenas, Edmund
TITLE OF INVENTION: An Enzyme Feed Additive and Animal Feed Including It
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESS: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/507,362A
FILING DATE: 27-Oct-1995
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Castaneda, Janet
REGISTRATION NUMBER: 33,228
REFERENCE/DOCKET NUMBER: GC226-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 846-4072
TELEFAX: (650) 845-6504
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 49 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-507-362A-19

Query Match      43.1%; Score 22.4; DB 4; Length 49;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY      1 ATGCAGTTCTCTCAGCCTCTCATGCTCTGCTGCGGCGGCTTGCC 48
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1 ATGAAGTTCCTTCAAGTCTCTCCCTGCTCATACCGGCGGCTTGCC 48

RESULT 33
US-09-146-770-2
Sequence 2, Application US/09146770
Patent No. 6187732
GENERAL INFORMATION:
APPLICANT: Fowler, Timothy
APPLICANT: Mutant Egit Cellulase, DNA Encoding
TITLE OF INVENTION: Such Egit Cellulase Compositions and Methods for Obtaining Same
FILE REFERENCE: GC546
CURRENT APPLICATION NUMBER: US/09/146,770
```

; CURRENT FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 702
; TYPE: DNA
; ORGANISM: T. longibrachiatum
US-09-146-770-2

Query Match 43.1%; Score 22.4; DB 3; Length 702;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 ATGCACTTCTTCACGCTCTCATCGCTGCTGCGCGGCGCTCGCC 48
Db 1 ATGAAGTTCCTTCAAGTCTCTCCCTGCGCTCATACCGCGCGCTCGCC 48

RESULT 34
US-09-216-295-2
; Sequence 2, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wentz, Dan J.
; TITLE OF INVENTION: No. 6268328e1 Variant EgIII-like Cellulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 702
; TYPE: DNA
; ORGANISM: Trichoderma longibrachiatum
US-09-216-295-2

Query Match 43.1%; Score 22.4; DB 3; Length 702;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 ATGCACTTCTTCACGCTCTCATCGCTGCTGCGCGGCGCTCGCC 48
Db 1 ATGAAGTTCCTTCAAGTCTCTCCCTGCGCTCATACCGCGCGCTCGCC 48

RESULT 35
US-09-633-084-2
; Sequence 2, Application US/09633084
; Patent No. 6407046
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; TITLE OF INVENTION: Mutant EgIII Cellulase, DNA Encoding
; TITLE OF INVENTION: Such EgIII Compositions and Methods for Obtaining Same
; FILE REFERENCE: GC546
; CURRENT APPLICATION NUMBER: US/09/633,084
; CURRENT FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 09/146,770
; PRIOR FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 702
; TYPE: DNA
; ORGANISM: T. longibrachiatum
US-09-633-084-2

Query Match 43.1%; Score 22.4; DB 4; Length 702;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 ATGCACTTCTTCACGCTCTCATCGCTGCTGCGCGGCGCTCGCC 48
Db 1 ATGAAGTTCCTTCAAGTCTCTCCCTGCGCTCATACCGCGCGCTCGCC 48

Query Match 43.1%; Score 22.4; DB 4; Length 702;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 1 ATGAAGTTCCTTCAAGTCTCTCCCTGCGCTCATACCGCGCGGCGCTCGCC 48

RESULT 36
US-10-075-872-2
; Sequence 2, Application US/10075872
; Patent No. 6500211
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; TITLE OF INVENTION: Mutant EgIII Cellulase, DNA Encoding
; TITLE OF INVENTION: Such EgIII Compositions and Methods for Obtaining Same
; FILE REFERENCE: GC546
; CURRENT APPLICATION NUMBER: US/10/075,872
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 702
; TYPE: DNA
; ORGANISM: T. longibrachiatum
US-10-075-872-2

Query Match 43.1%; Score 22.4; DB 4; Length 702;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 ATGCACTTCTTCACGCTCTCATCGCTGCTGCGCGGCGCTCGCC 48
Db 1 ATGAAGTTCCTTCAAGTCTCTCCCTGCGCTCATACCGCGCGGCGCTCGCC 48

RESULT 37
US-10-261-997-2
; Sequence 2, Application US/10261997
; Patent No. 6582750
; GENERAL INFORMATION:
; APPLICANT: Fowler, Timothy
; TITLE OF INVENTION: Mutant EgIII Cellulase, DNA Encoding
; TITLE OF INVENTION: Such EgIII Compositions and Methods for Obtaining Same
; FILE REFERENCE: GC546
; CURRENT APPLICATION NUMBER: US/10/261,997
; CURRENT FILING DATE: 2002-09-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 702
; TYPE: DNA
; ORGANISM: T. longibrachiatum
US-10-261-997-2

Query Match 43.1%; Score 22.4; DB 4; Length 702;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 ATGCACTTCTTCACGCTCTCATCGCTGCTGCGCGGCGCTCGCC 48
Db 1 ATGAAGTTCCTTCAAGTCTCTCCCTGCGCTCATACCGCGCGGCGCTCGCC 48

RESULT 38
US-08-032-848C-11
; Sequence 11, Application US/08032848C
; Patent No. 5473101
; GENERAL INFORMATION:
; APPLICANT: Ward, Michael
; APPLICANT: Clarkson, Kathleen A.
; APPLICANT: Weis, Geoffrey L.
; APPLICANT: Larens, Edward
; APPLICANT: Lorch, Jeffrey D.
; TITLE OF INVENTION: Purification and Molecular Cloning of
; TITLE OF INVENTION: EG III Cellulase
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Genencor International
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/032,848C
FILING DATE: MAR 17 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Horn, Margaret A.
REGISTRATION NUMBER: 33,401
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415 742-7356
TELEFAX: 415 742-7217
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1050 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-032-848C-11

Query Match 43.1%; Score 22.4; DB 1; Length 1050;
Best Local Similarity 66.7%; Pred. No. 54;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 ATGCACTCTCTCAGCGCTCTCATCGCTCTGCTGCGCGGCGCTCGCC 48
DB 151 ATGAATTCCTTCAAGTCTCTCCCTGCGCTCATACCGCGCGCTCGCC 198

RESULT 39
US-08-438-870-11
Sequence 11, Application US/08438870
Patent No. 5753484
GENERAL INFORMATION:
APPLICANT: Ward, Michael
APPLICANT: Clarkson, Kathleen A.
APPLICANT: Weiss, Geoffrey L.
APPLICANT: Larenas, Edward
APPLICANT: Lorch, Jeffrey D.
TITLE OF INVENTION: Purification and Molecular Cloning of EG
TITLE OF INVENTION: 111 Cellulase
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 180 Kimball Way
CITY: South San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/438,870
FILING DATE: May 10, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 35,696
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415 742-7555

TELEFAX: 415 742-7217
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1050 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-438-870-11

Query Match 43.1%; Score 22.4; DB 1; Length 1050;
Best Local Similarity 66.7%; Pred. No. 54;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 ATGCACTCTCTCAGCGCTCTCATCGCTCTGCTGCGCGGCGCTCGCC 48
DB 151 ATGAATTCCTTCAAGTCTCTCCCTGCGCTCATACCGCGCGCTCGCC 198

RESULT 40
US-08-774-065-1
Sequence 1, Application US/08774065
Patent No. 5969899
GENERAL INFORMATION:
APPLICANT: Bower, Benjamin
APPLICANT: Clarkson, Kathleen
APPLICANT: Larenas, Edmund
TITLE OF INVENTION: NOVEL OVERSIZED CELLULASE COMPOSITIONS
TITLE OF INVENTION: FOR USE IN DEREGMENT COMPOSITIONS AND
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENENCOR INTERNATIONAL
STREET: 925 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/774,065
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Glaister, Debra J.
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC368
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-846-7620
TELEFAX: 415-845-6504
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1050 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-774-065-1

Query Match 43.1%; Score 22.4; DB 2; Length 1050;
Best Local Similarity 66.7%; Pred. No. 54;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 ATGCACTCTCTCAGCGCTCTCATCGCTCTGCTGCGCGGCGCTCGCC 48
DB 151 ATGAATTCCTTCAAGTCTCTCCCTGCGCTCATACCGCGCGCTCGCC 198

Fri Nov 21 10:36:28 2003

us-10-081-935-1.rn1

Page 17

Search completed: November 13, 2003, 08:53:19
Job time : 65 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 08:18:16 (Search time 216 Seconds
(without alignments)
786.790 Million cell updates/sec

Title: US-10-081-935-1

Perfect score: 52
Sequence: 1 atgcagctctcctcagctctc.....cgctgcgcgcgcgcagctg 52

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published_Applications_NA.*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	48.1	336	9	US-09-825-414-37
2	25	48.1	1920	9	US-09-815-242-7958
3	24.6	47.3	1566	9	US-09-804-551B-43
4	24.6	47.3	1566	14	US-10-270-333-149
5	24.6	47.3	3772	14	US-10-270-333-148
6	24.4	46.9	925	12	US-10-029-386-7131
7	24.4	46.9	912	14	US-10-156-761-6808
8	24.4	46.9	954	14	US-10-156-761-7129
9	24.4	46.9	9025608	14	US-10-156-761-1
10	22.8	45.8	609	10	US-09-738-626-3337
11	23.8	45.8	1455	14	US-10-156-761-2054
12	23.8	45.8	3309400	10	US-09-738-626-1
13	23.6	45.4	590	12	US-10-027-632-50396
14	23.6	45.4	590	12	US-10-027-632-50397
15	23.6	45.4	590	12	US-10-027-632-50398
16	23.6	45.4	590	12	US-10-027-632-50399

17	23.6	45.4	590	12	US-10-027-632-82973	Sequence 82973, A
18	23.6	45.4	590	12	US-10-027-632-82974	Sequence 82974, A
19	23.6	45.4	590	13	US-10-027-632-50396	Sequence 50396, A
20	23.6	45.4	590	13	US-10-027-632-50397	Sequence 50397, A
21	23.6	45.4	590	13	US-10-027-632-50398	Sequence 50398, A
22	23.6	45.4	590	13	US-10-027-632-82972	Sequence 82972, A
23	23.6	45.4	590	13	US-10-027-632-82973	Sequence 82973, A
24	23.6	45.4	590	13	US-10-027-632-82974	Sequence 82974, A
25	23.6	45.4	1136	14	US-10-198-846-12577	Sequence 12577, A
26	23.6	45.4	135638	12	US-10-314-657-1	Sequence 1, Appl1
27	23.4	45.0	1029	14	US-10-156-761-5574	Sequence 5574, Ap
28	23.4	45.0	2466	12	US-09-953-348-83	Sequence 83, Appl
29	23.4	45.0	2466	14	US-10-267-255-83	Sequence 83, Appl
30	23.4	45.0	18034	12	US-09-953-348-75	Sequence 75, Appl
31	23.4	45.0	18034	14	US-10-267-255-75	Sequence 75, Appl
32	23.4	45.0	9025608	14	US-10-156-761-1	Sequence 1, Appl1
33	23.4	44.6	271	10	US-09-878-574-12667	Sequence 223, App
34	23.2	44.6	450	9	US-09-667-550-223	Sequence 223, App
35	23.2	44.6	497	12	US-10-029-386-21891	Sequence 21891, A
36	23.2	44.6	597	12	US-10-029-386-8191	Sequence 8191, Ap
37	23	44.2	492	11	US-09-918-995-37972	Sequence 37972, A
38	23	44.2	939	14	US-10-156-761-4691	Sequence 4691, Ap
39	22.8	43.8	322	10	US-09-960-352-476	Sequence 476, App
40	22.8	43.8	379	10	US-09-960-352-11092	Sequence 11092, A
41	22.8	43.8	381	10	US-09-960-352-12293	Sequence 12293, A
42	22.8	43.8	2559	14	US-10-128-714-7006	Sequence 7006, Ap
43	22.8	43.8	2621	14	US-10-128-714-6006	Sequence 6006, Ap
44	22.8	43.8	3608	14	US-10-128-714-6	Sequence 6, Appl1
45	22.8	43.8	4621	14	US-10-128-714-5006	Sequence 5006, Ap

ALIGNMENTS

```
RESULT 1
US-09-825-414-37
; Sequence 37, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Colimer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-09-825-414-37

Query Match      48.1% Score 25; DB 9; Length 336;
Best Local Similarity 69.4%; Pred. No. 12;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY      4 CAGTTCCTCAGCCTCTCATCGCTCGTGCCTGCAGCTGAGT 52
Db      15a CATTGGCGTGGCGCTCAACCCCTTGGATGATGCCGCCGCGCATTTG 202

RESULT 2
US-09-815-242-7958/c
; Sequence 7958, Application US/09815242
```

```
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7958
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1920)
US-09-815-242-7958

Query Match          48.1%; Score 25; DB 9; Length 1920;
Best Local Similarity 69.4%; Pred. No. 12;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY      4  CAGTTCTCTCAGCGCTCTCATCGCTGCTGCGCGCGCTCGCCAGTG 52
Db      507  CACCTCGCTGACCTCTCTGCTGCTGCGCGCGCGCGCGCGCGCGCG 459

RESULT 3
US-09-804-551B-43/c
; Sequence 43, Application US/09804551B
; Patent No. US20020056151A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Aktiengesellschaft
; TITLE OF INVENTION: Receptors for peptides from insects
; FILE REFERENCE: Le A 34 394
; CURRENT APPLICATION NUMBER: US/09/804,551B
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: DE 100 13 618.4
; PRIOR FILING DATE: 2000-03-18
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1563)
US-09-804-551B-43

Query Match          47.3%; Score 24.6; DB 9; Length 1566;
```

```
Best Local Similarity 70.2%; Pred. No. 16;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      3  GCAGTTCTCTCAGCGCTCTCATCGCTGCTGCGCGCGCTCGCCCA 49
Db      646  GCACAGCGGTACGACCTCTCTGCTGCTGCGCGCGAGTCTCCACA 600

RESULT 4
US-10-270-333-149/c
; Sequence 149, Application US/10270333
; Publication No. US20030092124A1
; GENERAL INFORMATION:
; APPLICANT: Cravchik, Anibal
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS, AND USES
; FILE REFERENCE: CL000733CON
; CURRENT APPLICATION NUMBER: US/10/270,333
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 60/168,677
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/175,691
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/191,638
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149
; LENGTH: 1566
; TYPE: DNA
; ORGANISM: Drosophila
US-10-270-333-149

Query Match          47.3%; Score 24.6; DB 14; Length 1566;
Best Local Similarity 70.2%; Pred. No. 16;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      3  GCAGTTCTCTCAGCGCTCTCATCGCTGCTGCGCGCGCTCGCCCA 49
Db      646  GCACAGCGGTACGACCTCTCTGCTGCTGCGCGCGAGTCTCCACA 600

RESULT 5
US-10-270-333-148/c
; Sequence 148, Application US/10270333
; Publication No. US20030092124A1
; GENERAL INFORMATION:
; APPLICANT: Cravchik, Anibal
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS, AND USES
; FILE REFERENCE: CL000733CON
; CURRENT APPLICATION NUMBER: US/10/270,333
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 60/168,677
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/175,691
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/191,638
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148
; LENGTH: 3772
; TYPE: DNA
; ORGANISM: Drosophila
US-10-270-333-148

Query Match          47.3%; Score 24.6; DB 14; Length 3772;
Best Local Similarity 70.2%; Pred. No. 16;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

Qy 3 GCAGTTCTCTCAGCGCTCTCATGCTCTGCTGCTGCCGCGCTCGGCCA 49
Db 1701 GCACAGCGCGTACGACGACTCTCTCTCTCGCTCGAGTCTCTCCACA 1655

RESULT 6
US-10-029-386-7131/c

; Sequence 7131, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7131
; LENGTH: 525
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AI021938.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: SWISSPROT HIT: Q92833, EVALU 1.00e-53
; OTHER INFORMATION: EST HUMAN HIT: BF311359.1, EVALU 0.00e+00
; OTHER INFORMATION: NT HIT: g16162622, EVALU 0.00e+00
US-10-029-386-7131

Query Match 46.9%; Score 24.4; DB 12; Length 525;
Best Local Similarity 68.0%; Pred. No. 20;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 TGCAGTTCTCTCAGCGCTCTCATGCTCTGCTGCTGCCGCGCTCGGCAGT 51
Db 138 TGCAGTTCTCTCAGCGCTCTCATGCTCTGCTGCTGCCGCGCTCGGCAGT 89

RESULT 7
US-10-156-761-6808/c

; Sequence 6808, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6808
; LENGTH: 912
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(912)
US-10-156-761-6808

Query Match 46.9%; Score 24.4; DB 14; Length 912;
Best Local Similarity 68.0%; Pred. No. 19;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 3 GCAGTTCTCTCAGCGCTCTCATGCTCTGCTGCTGCCGCGCTCGGCAGT 52
Db 523 GCAGTGGCGCAGCGGCTCTGACGAGCTTGGCGGTCCGCGCTCGGCCTTG 474

RESULT 8
US-10-156-761-7129

; Sequence 7129, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7129
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(954)
US-10-156-761-7129

Query Match 46.9%; Score 24.4; DB 14; Length 954;
Best Local Similarity 68.0%; Pred. No. 19;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 ATGCAGTTCTCTCAGCGCTCTCATGCTCTGCTGCTGCCGCGCTCGGCAGT 50
Db 853 ATGCAGTTCTCTCAGCGCTCTCATGCTCTGCTGCTGCCGCGCTCGGCAGT 902

RESULT 9
US-10-156-761-1/c

; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA

```
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match          46.9%; Score 24.4; DB 14; Length 9025608;
Best Local Similarity 68.0%; Pred. No. 14;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY      3 GCAGTCTCTCAGCCTCTCATCGCTCTCGTCGCGCGGCTGGCCAGTG 52
DB      8154610 GCAGGTGGCGCAGCGCTCGACGACGTCGTCCGCGTCCGCGCTCTTG 8154561

RESULT 10
US-09-738-626-3337
; Sequence 3337, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3337
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3337

Query Match          45.8%; Score 23.8; DB 10; Length 609;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      2 TGCAGTTCTCTCAGCCTCTCATCGCTCTCGTCGCGCGGCTGGCCAGTG 52
DB      401 TTCAGATCGGATACGCCATCATCGTATCCGCCGAGAGGTCTGCGCGTG 451

RESULT 11
US-10-156-761-2054
; Sequence 2054, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
```

```
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2054
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1455)
US-10-156-761-2054

Query Match          45.8%; Score 23.8; DB 14; Length 1455;
Best Local Similarity 66.7%; Pred. No. 31;
Matches 34; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      1 ATGCAGTTCTCTCAGCCTCTCATCGCTCTCGTCGCGCGGCTGGCCAGTG 51
DB      1282 ATGCCGTTCCAGTACATCATCACCGCGGCGTGGCGGCGGCGGCTCTCACT 1332

RESULT 12
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match          45.8%; Score 23.8; DB 10; Length 3309400;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 34; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      2 TGCAGTTCTCTCAGCCTCTCATCGCTCTCGTCGCGCGGCTGGCCAGTG 52
DB      3218393 TTCAGATCGGATACGCCATCATCGTATCCGCCGAGAGGTCTGCGCGTG 3218443

RESULT 13
US-10-027-632-50396/c
; Sequence 50396, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```


SEQ ID NO 82972
LENGTH: 590
TYPE: DNA
ORGANISM: Human
US-10-027-632-82972

Query Match 45.4%; Score 23.6; DB 12; Length 590;
Best Local Similarity 76.3%; Pred. No. 38;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGCAGTCTCTCAGCGCTCTCATCGCTCTGTCGCTGCC 39
DB 327 TCGTGTCTTCTCAAGCCACACACTCTGTCCTTACC 364

RESULT 17
US-10-027-632-82973

Sequence 82973, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 82973

LENGTH: 590

TYPE: DNA

ORGANISM: Human

US-10-027-632-82973

Query Match 45.4%; Score 23.6; DB 12; Length 590;
Best Local Similarity 76.3%; Pred. No. 38;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGCAGTCTCTCAGCGCTCTCATCGCTCTGTCGCTGCC 39
DB 327 TCGTGTCTTCTCAAGCCACACACTCTGTCCTTACC 364

RESULT 18
US-10-027-632-82974

Sequence 82974, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 82974

LENGTH: 590

TYPE: DNA

ORGANISM: Human

US-10-027-632-82974

Query Match 45.4%; Score 23.6; DB 12; Length 590;
Best Local Similarity 76.3%; Pred. No. 38;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGCAGTCTCTCAGCGCTCTCATCGCTCTGTCGCTGCC 39
DB 327 TCGTGTCTTCTCAAGCCACACACTCTGTCCTTACC 364

RESULT 19
US-10-027-632-50396/C

Sequence 50396, Application US/10027632
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 50396

LENGTH: 590

TYPE: DNA

ORGANISM: Human

US-10-027-632-50396

Query Match 45.4%; Score 23.6; DB 13; Length 590;
Best Local Similarity 76.3%; Pred. No. 38;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGCAGTCTCTCAGCGCTCTCATCGCTCTGTCGCTGCC 39
DB 264 TCGTGTCTTCTCAAGCCACACACTCTGTCCTTACC 227

RESULT 20
US-10-027-632-50397/C
Sequence 50397, Application US/10027632
GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide


```
/ TITLE OF INVENTION: Polymorphisms in the Human Genome
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ PRIOR FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 50397
/ LENGTH: 590
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-50397
```

```
Query Match
Best Local Similarity 45.4%; Score 23.6; DB 13; Length 590;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

```
Qy
2 TGCAGTTCTTCACGCTCTCATCGCTCTCGTCGCTGCC 39
Db
264 TGCTGTTTCTCAGCCGACACACTCTCGTCCCTACC 227
```

```
RESULT 21
US-10-027-632-50398/c
/ Sequence 50398, Application US/10027632
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 50398
/ LENGTH: 590
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-50398
```

```
Query Match
Best Local Similarity 45.4%; Score 23.6; DB 13; Length 590;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

```
Qy
2 TGCAGTTCTTCACGCTCTCATCGCTCTCGTCGCTGCC 39
```

```
Db
264 TGCTGTTTCTCAGCCGACACACTCTCGTCCCTACC 227
```

```
RESULT 22
US-10-027-632-82972
/ Sequence 82972, Application US/10027632
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 82972
/ LENGTH: 590
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-82972
```

```
Query Match
Best Local Similarity 45.4%; Score 23.6; DB 13; Length 590;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

```
Qy
2 TGCAGTTCTTCACGCTCTCATCGCTCTCGTCGCTGCC 39
Db
327 TGCTGTTTCTCAGCCGACACACTCTCGTCCCTACC 364
```

```
RESULT 23
US-10-027-632-82973
/ Sequence 82973, Application US/10027632
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 82973
/ LENGTH: 590
/ TYPE: DNA
```

ORGANISM: Human
US-10-027-632-82973

Query Match 45.4%; Score 23.6; DB 13; Length 590;
Best Local Similarity 76.3%; Pred. No. 38;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGCAGTCTCTCAGCGCTCTCATCGCTCTCGTCGCTGCC 39
DB 327 TGCTGTTTCTCAAGCCACACACTCTCTGCTCCTACC 364

RESULT 24
US-10-027-632-82974
Sequence 82974, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIORITY FILING DATE: 2002-04-30
PRIORITY APPLICATION NUMBER: US 60/218,006
PRIORITY FILING DATE: 2000-07-12
PRIORITY APPLICATION NUMBER: US 60/198,676
PRIORITY FILING DATE: 2000-04-20
PRIORITY APPLICATION NUMBER: US 60/193,483
PRIORITY FILING DATE: 2000-03-29
PRIORITY APPLICATION NUMBER: US 60/185,218
PRIORITY FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/167,363
PRIORITY FILING DATE: 1999-11-23
PRIORITY APPLICATION NUMBER: US 60/156,358
PRIORITY FILING DATE: 1999-09-28
PRIORITY APPLICATION NUMBER: US 60/146,002
PRIORITY FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 82974
LENGTH: 590
TYPE: DNA
ORGANISM: Human
US-10-027-632-82974

Query Match 45.4%; Score 23.6; DB 13; Length 590;
Best Local Similarity 76.3%; Pred. No. 38;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TGCAGTCTCTCAGCGCTCTCATCGCTCTCGTCGCTGCC 39
DB 327 TGCTGTTTCTCAAGCCACACACTCTCTGCTCCTACC 364

RESULT 25
US-10-198-846-12577
Sequence 12577, Application US/10198846
Publication No. US20030099974A1
GENERAL INFORMATION:
APPLICANT: Li, James

APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinhilber, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
FILE REFERENCE: MRI-049
CURRENT APPLICATION NUMBER: US/10/198,846
PRIORITY FILING DATE: 2002-07-18
PRIORITY APPLICATION NUMBER: 60/306,220
PRIORITY FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 12577

LENGTH: 1136
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136
OTHER INFORMATION: n = A,T,C or G
US-10-198-846-12577

Query Match 45.4%; Score 23.6; DB 14; Length 1136;
Best Local Similarity 69.6%; Pred. No. 37;
Matches 32; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 ATGCAGTCTCTCAGCGCTCTCATCGCTCTCGTCGCTGCCGCTCG 46
DB 54 ATGCAGTCTCTCAGCGCTCTCATCGCTCTCGTCGCTGCCGCTCG 99

RESULT 26
US-10-314-657-1/C
Sequence 1, Application US/10314657
Publication No. US20030175888A1
GENERAL INFORMATION:
APPLICANT: SHEN, Ben

APPLICANT: TANG, Gong-Li
TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
FILE REFERENCE: 054030-0021
CURRENT APPLICATION NUMBER: US/10/314,657
PRIORITY FILING DATE: 2002-12-09
PRIORITY APPLICATION NUMBER: PCT/US02/08937
PRIORITY FILING DATE: 2002-03-22
PRIORITY APPLICATION NUMBER: US 60/278,935
PRIORITY FILING DATE: 2001-03-26
NUMBER OF SEQ ID NOS: 214
SOFTWARE: PatentIn Version 3.2
SEQ ID NO 1
LENGTH: 135638
TYPE: DNA
ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

Query Match 45.4%; Score 23.6; DB 12; Length 135638;
Best Local Similarity 76.3%; Pred. No. 32;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 15 CGCTCTATCGCTCTCATCGCTCTCGTCGCTGCCGCTCGCAGTG 52
DB 34257 CGCTCTATCGCTCTCATCGCTCTCGTCGCTGCCGCTCGCAGTG 34220

RESULT 27
US-10-156-761-5574
Sequence 5574, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
PRIORITY FILING DATE: 2002-05-29
PRIORITY APPLICATION NUMBER: JP 2001-204089
PRIORITY FILING DATE: 2001-05-30
PRIORITY APPLICATION NUMBER: JP 2001-272697
PRIORITY FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109

```

; SEQ ID NO 5574
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1029)
US-10-156-761-5574
```

```

Query Match          45.0%; Score 23.4; DB 14; Length 1029;
Best Local Similarity 67.3%; Pred. No. 43;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```

Qy      1  ATGCAGTTCTCTCAGCGCTTCATCGCTCTGCTGCGCGCGCTCGCCA 49
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      10  AAGCGTACGTCCTCCTCTGACCGCGCTGCTGCGCGCGCTCGCCA 58
```

RESULT 28

```

US-09-953-348-83
; Sequence 83, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqing
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US00/06394
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/266965
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 2466
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-953-348-83
```

```

Query Match          45.0%; Score 23.4; DB 12; Length 2466;
Best Local Similarity 73.2%; Pred. No. 42;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```

Qy      9  CTCACGCTCTCATCGCTCTGCTGCTGCGCGCGCTCGCCA 49
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      48  CGCTGCCCGCCTGATGCGCCTGCTGCTCGCGCACCGCCA 88
```

RESULT 29

```

US-10-267-255-83
; Sequence 83, Application US/10267255
; Publication No. US20030124689A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
```

```

; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 83
; LENGTH: 2466
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-10-267-255-83
```

```

Query Match          45.0%; Score 23.4; DB 14; Length 2466;
Best Local Similarity 73.2%; Pred. No. 42;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```

Qy      9  CTCACGCTCTCATCGCTCTGCTGCTGCGCGCGCTCGCCA 49
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      48  CGCTGCCCGCCTGATGCGCCTGCTGCTCGCGCACCGCCA 88
```

RESULT 30

```

US-09-953-348-75
; Sequence 75, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingqing
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US00/06394
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/266965
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 75
; LENGTH: 18034
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (302)..(302)
; OTHER INFORMATION: n is a or t or g or c
US-09-953-348-75
```

```

Query Match          45.0%; Score 23.4; DB 12; Length 18034;
Best Local Similarity 73.2%; Pred. No. 40;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```

Qy      9  CTCACGCTCTCATCGCTCTGCTGCTGCGCGCGCTCGCCA 49
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      11562 CGCTGCCCGCCTGATGCGCCTGCTGCTCGCGCACCGCCA 11602
```

RESULT 31

```

US-10-267-255-75
; Sequence 75, Application US/10267255
; Publication No. US20030124689A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
```

```

; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 75
; LENGTH: 18034
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (302)...(302)
; OTHER INFORMATION: n is a or t or g or c
US-10-267-255-75

```

```

Query Match          45.0%; Score 23.4; DB 14; Length 18034;
Best Local Similarity 73.2%; Pred. No. 40;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```

```

QY 9 CTCTACGCTCTCATGCGCTCTGTCGCGCGGCTTGCCA 49
DB 11562 CGCTGCGCGCCCTGATGCGCTCTGTCGCGGACCGCA 11602

```

```

RESULT 32
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermectilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

```

Query Match          45.0%; Score 23.4; DB 14; Length 9025608;
Best Local Similarity 67.3%; Pred. No. 31;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

```

QY 1 ATGCAGTTCCTCAGCGCTCTCATGCGCTCTGTCGCGCGGCTTGCCA 49
DB 6765271 AAGCGTAGCGTCTCTCTGACGCGGCTCTGCGCGGCTTGCCA 6765319

```

```

RESULT 33
US-09-878-574-12667/c
; Sequence 12667, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21 (15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 12667
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(271)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 70106061H1
US-09-878-574-12667

```

```

Query Match          44.6%; Score 23.2; DB 10; Length 271;
Best Local Similarity 70.5%; Pred. No. 53;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 8 TCTTCACGCTCTCATGCGCTCTGTCGCGCGGCTTGCGCAGT 51
DB 59 TCTGTCCTCTCTCATGCTCATCTCATCTTGTGGCCACGCGCAGT 16

```

```

RESULT 34
US-09-867-550-223/c
; Sequence 223, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells an
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: US96/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 223
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(31)
; OTHER INFORMATION: wherein any n is one of a or t or g or c
US-09-867-550-223

```

```

Query Match          44.6%; Score 23.2; DB 9; Length 450;
Best Local Similarity 77.8%; Pred. No. 52;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 13 CAGGCTCTCATGCGCTCTGTCGCGCGGCTTGCGCC 48
DB 207 CTCGATCTCGCGGCTCTGTCGCGGCTTGCGCC 172

```

```

RESULT 35
US-10-029-386-21891
; Sequence 21891, Application US/10029386
; Publication No. US20030194704A1

```


Db 854 TGCAGCGCGCCCGCCCTGGACGGCCTGGTCCGCCGACGGCCTCGTC 900

RESULT 39
US-09-960-352-476

; Sequence 476, Application US/09960352

Patent No. US20020137139A1

GENERAL INFORMATION

APPLICANT: Tao, Ne

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, N

TITLE OF INVENTION: NUCLEI

FILE OF INVENTION: MUSCLE AND F
REFERENCE 16511 006/37-21(1

CURRENT APPLICATION NUMBER: US/09/960.352

CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 476

TYPE: DNA
LENGTH: 322
;

ORGANISM:

OTHER INFORMATION:

US-09-960-352-476

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

US-09-960-352-476

Query Match	43.8%	Score 22.8;	DB 10;	Length 322;
Best Local Similarity	71.4%	Pred. No. 73;		
Matches 30;	Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;

```

QY      8 TCTCTACGCTCTCATCGCTCTGTCGTCGCGGCGCTCGCA 49
      | | | | | | | | | | | | | | | | | | | |
Db    281 TATGCGCGCTCTTTCGCTTTCGTCCTTCCTTCTCGCA 322

```

RESULT 40
IIS-09-060

US-03-380-332-11032
: Sequence 11092 A

Patent No. US20020137139A1

; GENERAL INFORMATION:

APPLICANT: Warren,

APPLICANT: Tao, Nengbing

APPLICANT: Mathialagan, N

TITLE OF INVENTION: NUCLEIC ACID

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960,352

NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 11092

; LENGTH: 379

```

; TYPE: DNA

```

ORGANISM:

OTHER INFORMATION: C
HS-09-960-352-111092

C
C
I
C
C
I
U
C
C
L
L
P
I
H
H
C
C
N

US-09-960-352-11092

Query Match	43.8%	Score 22.8;	DB 10;	Length 379;
Best Local Similarity	71.4%	Pred. No. 73;		
Matches 30;	Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;

QY 8 TCTCTACAGCTTCATCGTCTGTGCTGCCGGCCTCGCA 49
| | | | | | | | | | | | | | | | | | | |
Db 281 TATCGCCGCCTCCTTGCTTTCTTGCTTCTTTCTCGCA 322

Search completed: November 13, 2003, 09:48:44
Job time : 244 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 08:54:45 (without alignments)

1080.396 Million cell updates/sec

Title: US-10-081-935-1

Perfect score: 52

Sequence: 1 atgcagctctctcagctctc.....cgctgccgcgcctcgacgtg 52

Scoring table: OLIGO-NUC

Gapop_60.0, Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database:

GenEmbl:.*
1: gb_ba:.*
2: gb_hcg:.*
3: gb_in:.*
4: gb_cm:.*
5: gb_ov:.*
6: gb_pac:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_scs:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vi:.*
15: em_ba:.*
16: em_fun:.*
17: em_hum:.*
18: em_in:.*
19: em_mu:.*
20: em_om:.*
21: em_ov:.*
22: em_ov:.*
23: em_pac:.*
24: em_ph:.*
25: em_pl:.*
26: em_ro:.*
27: em_scs:.*
28: em_un:.*
29: em_vi:.*
30: em_hcg_hum:.*
31: em_hcg_inv:.*
32: em_hcg_other:.*
33: em_hcg_mus:.*
34: em_hcg_pln:.*
35: em_hcg_rtd:.*
36: em_hcg_mam:.*
37: em_hcg_vtc:.*
38: em_sy:.*
39: em_hcgo_hum:.*
40: em_hcgo_mus:.*
41: em_hcgo_other:.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	100.0	1201	8	CIU39835
2	52	100.0	1234	8	CIU32518
3	52	100.0	1435	8	CIU51200
4	52	100.0	3821	8	AF013256
5	52	100.0	3821	8	AF013256
6	52	100.0	3821	8	AF013256
7	52	100.0	3821	8	AF013256
8	52	100.0	3821	8	AF013256
9	52	100.0	3821	8	AF013256
10	52	100.0	3821	8	AF013256
11	52	100.0	3821	8	AF013256
12	52	100.0	3821	8	AF013256
13	52	100.0	3821	8	AF013256
14	52	100.0	3821	8	AF013256
15	52	100.0	3821	8	AF013256
16	52	100.0	3821	8	AF013256
17	52	100.0	3821	8	AF013256
18	52	100.0	3821	8	AF013256
19	52	100.0	3821	8	AF013256
20	52	100.0	3821	8	AF013256
21	52	100.0	3821	8	AF013256
22	52	100.0	3821	8	AF013256
23	52	100.0	3821	8	AF013256
24	52	100.0	3821	8	AF013256
25	52	100.0	3821	8	AF013256
26	52	100.0	3821	8	AF013256
27	52	100.0	3821	8	AF013256
28	52	100.0	3821	8	AF013256
29	52	100.0	3821	8	AF013256
30	52	100.0	3821	8	AF013256
31	52	100.0	3821	8	AF013256
32	52	100.0	3821	8	AF013256
33	52	100.0	3821	8	AF013256
34	52	100.0	3821	8	AF013256
35	52	100.0	3821	8	AF013256
36	52	100.0	3821	8	AF013256
37	52	100.0	3821	8	AF013256
38	52	100.0	3821	8	AF013256
39	52	100.0	3821	8	AF013256
40	52	100.0	3821	8	AF013256
41	52	100.0	3821	8	AF013256
42	52	100.0	3821	8	AF013256
43	52	100.0	3821	8	AF013256
44	52	100.0	3821	8	AF013256
45	52	100.0	3821	8	AF013256
46	52	100.0	3821	8	AF013256
47	52	100.0	3821	8	AF013256
48	52	100.0	3821	8	AF013256
49	52	100.0	3821	8	AF013256
50	52	100.0	3821	8	AF013256
51	52	100.0	3821	8	AF013256
52	52	100.0	3821	8	AF013256
53	52	100.0	3821	8	AF013256
54	52	100.0	3821	8	AF013256
55	52	100.0	3821	8	AF013256
56	52	100.0	3821	8	AF013256
57	52	100.0	3821	8	AF013256
58	52	100.0	3821	8	AF013256
59	52	100.0	3821	8	AF013256
60	52	100.0	3821	8	AF013256
61	52	100.0	3821	8	AF013256
62	52	100.0	3821	8	AF013256
63	52	100.0	3821	8	AF013256
64	52	100.0	3821	8	AF013256
65	52	100.0	3821	8	AF013256

C 66	17	32.7	191022	8	AP003431	AP003431 Oryza sat	C 139	16	30.8	139468	8	AC090441	AC090441 Oryza sat
C 67	17	32.7	200050	1	AL646068	AL646068 Ralscomia	C 140	16	30.8	143618	2	AC026855	AC026855 Homo sapi
C 68	17	32.7	220676	2	AC114574	AC114574 Mus muscu	C 141	16	30.8	150716	8	AC137928	AC137928 Oryza sat
C 69	17	32.7	231461	1	AC096278	AC096278 Rattus no	C 142	16	30.8	152347	2	AC123898	AC123898 Medicago
C 70	17	32.7	244757	3	AE003678	AE003678 Drosophi	C 143	16	30.8	152691	2	AC141704	AC141704 Apis mell
C 71	17	32.7	247590	2	AC119964	AC119964 Mus muscu	C 144	16	30.8	153344	2	AC104160	AC104160 Canis fam
C 72	17	32.7	252602	2	AC091704	AC091704 Mus muscu	C 145	16	30.8	153438	2	AP005907	AP005907 Oryza sat
C 73	17	32.7	263668	2	AC126697	AC126697 Rattus no	C 146	16	30.8	157006	2	AP003860	AP003860 Oryza sat
C 74	17	32.7	299850	1	AP005949	AP005949 Bradyrhiz	C 147	16	30.8	159029	2	AP004692	AP004692 Oryza sat
C 75	16	30.8	550	8	HS432669	HS432669 Homo sapi	C 148	16	30.8	160356	9	AC009068	AC009068 Homo sapi
C 76	16	30.8	610	8	AF297656	AF297656 Pelotum	C 149	16	30.8	160743	9	HS4258F22	HS4258F22 Homo sapi
C 77	16	30.8	664	3	DME426941	DME426941 Drosophi	C 150	16	30.8	162339	8	AC022457	AC022457 Oryza sat
C 78	16	30.8	909	1	PSEWAB	M24016 P. mevaloni	C 151	16	30.8	163178	2	AC008336	AC008336 Drosophi
C 79	16	30.8	1117	1	PSEWAB	M31807 P. mevaloni	C 152	16	30.8	163450	8	OSJ000127	OSJ000127 Oryza sat
C 80	16	30.8	1286	8	BC000207	BC000207 Homo sapi	C 153	16	30.8	167807	2	AP004001	AP004001 Oryza sat
C 81	16	30.8	1307	9	BC014040	BC014040 Homo sapi	C 154	16	30.8	169006	2	AC142178	AC142178 Rattus no
C 82	16	30.8	1653	6	AX358336	AX358336 Sequence	C 155	16	30.8	169931	2	AC008321	AC008321 Homo sapi
C 83	16	30.8	1705	1	AB029336	AB029336 BurKhole	C 156	16	30.8	172378	2	BX470161	BX470161 Danio rer
C 84	16	30.8	1871	9	BC043607	BC043607 Homo sapi	C 157	16	30.8	174583	10	AL723313	AL723313 Mouse DNA
C 85	16	30.8	1889	9	BC027475	BC027475 Homo sapi	C 158	16	30.8	176872	10	AL772313	AL772313 Mouse DNA
C 86	16	30.8	1893	3	AF403542	AF403542 Bombyx mo	C 159	16	30.8	177163	9	AC100786	AC100786 Homo sapi
C 87	16	30.8	2248	3	DROEX02	L07660 Drosophila	C 160	16	30.8	177472	9	AC008786	AC008786 Homo sapi
C 88	16	30.8	2484	9	BC015581	BC015581 Homo sapi	C 161	16	30.8	179369	2	AC122291	AC122291 Rattus no
C 89	16	30.8	2484	9	BC007982	BC007982 Homo sapi	C 162	16	30.8	180153	2	AC018798	AC018798 Homo sapi
C 90	16	30.8	2604	3	AF419623	AF419623 Drosophi	C 163	16	30.8	181401	2	AL954127	AL954127 Mus muscu
C 91	16	30.8	2760	1	SRENSRS	Y14336 Streptomyce	C 164	16	30.8	182769	9	AP003176	AP003176 Homo sapi
C 92	16	30.8	3040	3	AF419622	AF419622 Drosophi	C 165	16	30.8	183065	2	AC008401	AC008401 Homo sapi
C 93	16	30.8	3750	1	U93704	U93704 Riffia pach	C 166	16	30.8	183426	2	AC107196	AC107196 Rattus no
C 94	16	30.8	4390	9	BC037556	BC037556 Homo sapi	C 167	16	30.8	184554	3	AC099022	AC099022 Drosophi
C 95	16	30.8	5178	3	AY119603	AY119603 Drosophi	C 168	16	30.8	185200	3	AC092401	AC092401 Drosophi
C 96	16	30.8	7133	9	D80004	D80004 Homo sapien	C 169	16	30.8	185290	2	AC107006	AC107006 Rattus no
C 97	16	30.8	9787	1	AE011672	AE011672 Xanthomon	C 170	16	30.8	185321	9	AC123908	AC123908 Homo sapi
C 98	16	30.8	10049	1	AE014662	AE014662 Bifidobac	C 171	16	30.8	188624	10	AL772282	AL772282 Mouse DNA
C 99	16	30.8	10417	1	AE004546	AE004546 Pseudomoc	C 172	16	30.8	188646	9	AC027045	AC027045 Homo sapi
C 100	16	30.8	10444	14	PV1422133	PV1422133 Pseudorab	C 173	16	30.8	189252	2	AC018930	AC018930 Homo sapi
C 101	16	30.8	10542	1	AE001951	AE001951 Deinococ	C 174	16	30.8	190187	2	AC121140	AC121140 Mus muscu
C 102	16	30.8	10600	1	RM106EXB	RM106EXB Rhizobium	C 175	16	30.8	190615	2	AC108411	AC108411 Mus muscu
C 103	16	30.8	10637	1	AE004805	AE004805 Pseudomon	C 176	16	30.8	191162	3	AC008865	AC008865 Homo sapi
C 104	16	30.8	10967	1	AE007262	AE007262 Stenohizo	C 177	16	30.8	191558	3	AC013431	AC013431 Drosophi
C 105	16	30.8	11744	1	AE005841	AE005841 Caulobac	C 178	16	30.8	193367	1	AE001977	AE001977 Homo sapi
C 106	16	30.8	13947	4	AY102170	AY102170 Canis fam	C 179	16	30.8	198677	2	AC001863	AC001863 Drosophi
C 107	16	30.8	15168	5	PSNARXL	Y15552 Pseudomonas	C 180	16	30.8	198764	9	AC103809	AC103809 Homo sapi
C 108	16	30.8	15670	5	AF090339	AF090339 Rhea amer	C 181	16	30.8	200900	9	AC013410	AC013410 Homo sapi
C 109	16	30.8	25108	2	AC140855	AC140855 Botryllus	C 182	16	30.8	201785	2	AC141171	AC141171 Rattus no
C 110	16	30.8	27685	2	AC019935	AC019935 Drosophi	C 183	16	30.8	202190	2	AC026813	AC026813 Mus muscu
C 111	16	30.8	39089	4	AF483210	AF483210 Canis fam	C 184	16	30.8	204712	2	AC134627	AC134627 Rattus no
C 112	16	30.8	39516	3	AC058781	AC058781 Leishmani	C 185	16	30.8	206387	2	AC130934	AC130934 Rattus no
C 113	16	30.8	44806	9	AC069162	AC069162 Homo sapi	C 186	16	30.8	209740	10	AL772166	AL772166 Mouse DNA
C 114	16	30.8	53339	2	AC091847	AC091847 Homo sapi	C 187	16	30.8	214416	2	AC112406	AC112406 Rattus no
C 115	16	30.8	58343	1	AB089954	AB089954 Micromono	C 188	16	30.8	216807	2	AC124146	AC124146 Rattus no
C 116	16	30.8	61174	2	AC080104	AC080104 Homo sapi	C 189	16	30.8	219694	2	AC111438	AC111438 Rattus no
C 117	16	30.8	72899	2	AC139009	AC139009 Homo sapi	C 190	16	30.8	226874	2	AC099458	AC099458 Homo sapi
C 118	16	30.8	74821	9	AC109642	AC109642 Homo sapi	C 191	16	30.8	229560	2	BX005404	BX005404 Danio rer
C 119	16	30.8	77137	3	AC004422	AC004422 Drosophi	C 192	16	30.8	234133	2	AC095750	AC095750 Rattus no
C 120	16	30.8	80139	8	AC027135	AC027135 Arabidops	C 193	16	30.8	239781	2	AC106535	AC106535 Rattus no
C 121	16	30.8	83876	2	AC020009	AC020009 Drosophi	C 194	16	30.8	245210	2	AC137771	AC137771 Homo sapi
C 122	16	30.8	84400	8	AB010075	AB010075 Arabidops	C 195	16	30.8	246792	2	BX470224	BX470224 Danio rer
C 123	16	30.8	91200	8	AP005816	AP005816 Oryza sat	C 196	16	30.8	248091	2	AC131164	AC131164 Rattus no
C 124	16	30.8	96035	2	AC105319	AC105319 Oryza sat	C 197	16	30.8	248366	2	AC096062	AC096062 Rattus no
C 125	16	30.8	104170	2	AC091510	AC091510 Leishmani	C 198	16	30.8	258950	2	AC099160	AC099160 Rattus no
C 126	16	30.8	110000	2	AC113913	AC113913 Oryza sat	C 199	16	30.8	264629	2	AC095776	AC095776 Rattus no
C 127	16	30.8	110000	2	LMFLCHR13_13	Continuation (15 of	C 200	16	30.8	276881	1	AC095481	AC095481 Rattus no
C 128	16	30.8	110000	2	LMFLCHR12_14	Continuation (15 of	C 201	16	30.8	290714	1	AE017025	AE017025 Bacillus
C 129	16	30.8	110000	2	LMFLCHR31_20	Continuation (21 of	C 202	16	30.8	295500	1	AP005954	AP005954 Bradyrhiz
C 130	16	30.8	110000	2	LMFLCHR36_03	Continuation (4 of	C 203	16	30.8	295775	2	AC096061	AC096061 Rattus no
C 131	16	30.8	118086	8	ATF6H11	AL021684 Arabidops	C 204	16	30.8	297470	2	AC125817	AC125817 Rattus no
C 132	16	30.8	127545	2	BX324148	BX324148 Danio rer	C 205	16	30.8	299850	1	AP005963	AP005963 Bradyrhiz
C 133	16	30.8	129430	9	AC008905	AC008905 Homo sapi	C 206	16	30.8	300029	8	AE017103	AE017103 Oryza sat
C 134	16	30.8	133353	2	AC131246	AC131246 Oryza sat	C 207	16	30.8	301250	3	AE005950	AE005950 Drosophi
C 135	16	30.8	135809	8	AC079021	AC079021 Oryza sat	C 208	16	30.8	303345	1	AE003500	AE003500 Rattus no
C 136	16	30.8	135876	8	AC113337	AC113337 Genomc s	C 209	16	30.8	306050	8	RME603645	RME603645 Rhizobium
C 137	16	30.8	138335	8	AC119148	AC119148 Oryza sat	C 210	16	30.8	306057	8	AE017063	AE017063 Oryza sat
C 138	16	30.8	139072	9	AP000485	AP000485 Homo sapi	C 211	16	30.8	307053	8	AE017079	AE017079 Oryza sat

C 212	16	30.8	314957	3	AE003581	Drosophila
C 213	16	30.8	346294	1	AP002999	Mesorhizo
C 214	16	30.8	349498	1	AP003002	Mesorhizo
C 215	16	30.8	349980	6	AX493782	Sequence
C 216	16	30.8	349980	6	AX553949	Sequence
C 217	16	28.8	226	8	AT020310	Oryza sat
C 218	15	28.8	260	9	HS71E3R	H. sapiens C
C 219	15	28.8	272	14	BLVNADMA	119257 Bovine Leuk
C 220	15	28.8	329	14	ASSVCG	X17696 Apple scar
C 221	15	28.8	330	3	AY165174	AY165174 Tomcus p
C 222	15	28.8	330	14	ASSVXK	AY0435 Apple scar
C 223	15	28.8	331	14	AF421195	AF421195 Apple sca
C 224	15	28.8	331	14	DAVDCG	X71599 Dapple Appl
C 225	15	28.8	423	16	AR233241	AR233241 Sequence
C 226	15	28.8	444	1	UBA306168	AJ06168 unculture
C 227	15	28.8	448	6	AR128198	AR128198 Sequence
C 228	15	28.8	471	6	BD025746	BD025746 Sequence
C 229	15	28.8	535	11	G83520	G83520 S208P674FH
C 230	15	28.8	536	12	AY199102	AY199102 Arabidops
C 231	15	28.8	571	8	ZW2EM2	X85335 Z. mays mRNA
C 232	15	28.8	571	11	BV016835	BV016835 S208P6272
C 233	15	28.8	667	6	BD147972	BD147972 Primer fo
C 234	15	28.8	671	9	HS4331133	AJ331133 Homo sapi
C 235	15	28.8	680	8	MZEMADSC	L46399 Zea mays MA
C 236	15	28.8	694	3	AF438233	AF438233 Leishmani
C 237	15	28.8	753	6	BD150581	BD150581 Primer fo
C 238	15	28.8	859	4	RABKCHVG	L01789 Oryzicolaus
C 239	15	28.8	886	8	ZW2EM1	X85334 Z. mays mRNA
C 240	15	28.8	903	8	ZW2EM3	X85336 Z. mays mRNA
C 241	15	28.8	927	9	BC017347	BC017347 Homo sapi
C 242	15	28.8	1032	6	AX653040	AX653040 Sequence
C 243	15	28.8	1053	1	UBA494772	AJ94772 Unculture
C 244	15	28.8	1054	10	BC021770	BC021770 Mus muscu
C 245	15	28.8	1113	3	LME313202	AJ13302 Leishmani
C 246	15	28.8	1115	5	CHKMRP4	AJ1059 Chicken MRF
C 247	15	28.8	1190	6	AX654314	AX654314 Sequence
C 248	15	28.8	1201	8	AF264731	AF264731 Oryza sat
C 249	15	28.8	1314	8	AF327528	AF327528 Arabidops
C 250	15	28.8	1335	8	AF324719	AF324719 Arabidops
C 251	15	28.8	1351	5	AF232670	AF232670 Protocpter
C 252	15	28.8	1374	1	HS222E131	HL45014 Novel hum
C 253	15	28.8	1491	1	SFU09533	U90533 Streptomyc
C 254	15	28.8	1561	8	AY056192	AY056192 Arabidops
C 255	15	28.8	1597	8	AF207593	AF207593 Arabidops
C 256	15	28.8	1724	5	CJUI2641	UI2641 Coturnix ja
C 257	15	28.8	1769	9	BC004290	BC004290 Homo sapi
C 258	15	28.8	1777	3	AK132664	AK132664 Ciona int
C 259	15	28.8	2000	6	AX508283	AX508283 Sequence
C 260	15	28.8	2024	10	BC027051	BC027051 Mus muscu
C 261	15	28.8	2034	8	AB080193	AB080193 Pisum sat
C 262	15	28.8	2078	9	BC033099	BC033099 Homo sapi
C 263	15	28.8	2112	9	BC041802	BC041802 Homo sapi
C 264	15	28.8	2184	9	BC030005	BC030005 Homo sapi
C 265	15	28.8	2212	8	AY140043	AY140043 Arabidops
C 266	15	28.8	2336	3	AY071759	AY071759 Drosophila
C 267	15	28.8	2467	5	PCU44726	U44726 Penicillium
C 268	15	28.8	2494	5	AF011356	AF011356 Gallus ga
C 269	15	28.8	2520	9	AF206328	AF206328 Homo sapi
C 270	15	28.8	2554	1	BBR303058	AJ303058 Bordetell
C 271	15	28.8	2554	1	BBR303059	AJ303059 Bordetell
C 272	15	28.8	2554	1	BBR303060	AJ303060 Bordetell
C 273	15	28.8	2554	1	BBR303061	AJ303061 Bordetell
C 274	15	28.8	2554	1	BBR303062	AJ303062 Bordetell
C 275	15	28.8	2554	1	BBR303063	AJ303063 Bordetell
C 276	15	28.8	2554	1	BBR303064	AJ303064 Bordetell
C 277	15	28.8	2554	1	BBR303066	AJ303066 Bordetell
C 278	15	28.8	2554	1	BBR303071	AJ303071 Bordetell
C 279	15	28.8	2556	6	BD159688	BD159688 Primer fo
C 280	15	28.8	2556	6	AK022879	AK022879 Homo sapi
C 281	15	28.8	2580	6	BC008766	BC008766 Homo sapi
C 282	15	28.8	2598	6	AX714945	AX714945 Sequence
C 283	15	28.8	2598	9	AK057694	AK057694 Homo sapi
C 284	15	28.8	2617	6	BD158744	BD158744 Primer fo
C 285	15	28.8	2617	9	AK023976	AK023976 Homo sapi
C 286	15	28.8	2650	3	AF144322	AF144322 Giardia i
C 287	15	28.8	2693	9	BC014906	BC014906 Homo sapi
C 288	15	28.8	2697	9	BC044887	BC044887 Homo sapi
C 289	15	28.8	2702	9	AF203474	AF203474 Homo sapi
C 290	15	28.8	2790	6	AX333036	AX333036 Sequence
C 291	15	28.8	2790	6	AX335818	AX335818 Sequence
C 292	15	28.8	2790	6	HSU10868	U10868 Human aldeh
C 293	15	28.8	2794	8	BT004338	BT004338 Arabidops
C 294	15	28.8	2808	9	BC001655	BC001655 Homo sapi
C 295	15	28.8	2808	9	BC001843	BC001843 Homo sapi
C 296	15	28.8	2821	8	AF443119	AF443119 Usillaago
C 297	15	28.8	2837	9	BC013584	BC013584 Homo sapi
C 298	15	28.8	2947	1	AF239164	AF239164 Rhizobiu
C 299	15	28.8	2964	9	AK058012	AK058012 Homo sapi
C 300	15	28.8	2969	8	AY059771	AY059771 Arabidops
C 301	15	28.8	3391	1	TFALAS	X95571 T. ferrooxid
C 302	15	28.8	3514	3	DWSING3	X17550 D. melanoga
C 303	15	28.8	3866	3	BT003619	BT003619 Drosophila
C 304	15	28.8	3963	6	AX260206	AX260206 Sequence
C 305	15	28.8	4038	1	MXALON	D12923 Myxococcu
C 306	15	28.8	4198	6	AX067966	AX067966 Sequence
C 307	15	28.8	4198	8	ATENRA	Y13860 Arabidopsis
C 308	15	28.8	4359	3	AY094956	AY094956 Drosophila
C 309	15	28.8	4463	8	AF355592	AF355592 Filobaasid
C 310	15	28.8	4475	8	ZW2EMAGEN	X91882 Z. mays ZEMA
C 311	15	28.8	4516	9	AB037841	AB037841 Homo sapi
C 312	15	28.8	4750	3	AF197953	AF197953 Topoplaem
C 313	15	28.8	4800	1	AB044351	AB044351 Pseudomon
C 314	15	28.8	4842	3	AF132179	AF132179 Drosophila
C 315	15	28.8	4945	10	RN0550165	AJ550165 Rattus no
C 316	15	28.8	4986	2	AC019465	AC019465 Drosophila
C 317	15	28.8	5118	6	A79572	A79572 Sequence 3
C 318	15	28.8	5118	6	NR175232	NR175232 Sequence
C 319	15	28.8	5170	10	RN0426052	AJ26052 Rattus no
C 320	15	28.8	5334	1	BBCYAG	Z7112 Bordetella
C 321	15	28.8	5369	1	BPA249835	AJ249835 Bordetella
C 322	15	28.8	5419	3	DME238847	AJ38847 Drosophila
C 323	15	28.8	5694	14	HSCZCHOM	UJ2062 Cercopithec
C 324	15	28.8	5721	6	A25910	A25910 B. bronchiet
C 325	15	28.8	5962	1	MBS100PER	Y13228 Mycobacteri
C 326	15	28.8	6000	6	AF433876	AF433876 Giardia i
C 327	15	28.8	6059	6	AX281301	AX281301 Sequence
C 328	15	28.8	6078	3	AB061681	AB061681 Drosophila
C 329	15	28.8	6441	6	A14850	A14850 B. pertussis
C 330	15	28.8	6441	6	A79570	A79570 Sequence 1
C 331	15	28.8	6441	6	AR175231	AR175231 Sequence
C 332	15	28.8	6442	1	BPCYA	Y00545 Bordetella
C 333	15	28.8	6443	6	A07292	A07292 B. pertussis
C 334	15	28.8	6443	6	A32503	A32503 B. anthracis
C 335	15	28.8	6991	8	AB084921	AB084921 Paecilomy
C 336	15	28.8	7646	1	AF116906	AF116906 Rhodococc
C 337	15	28.8	9349	1	AF173961	AF173961 Thauera a
C 338	15	28.8	9405	6	AX399934	AX399934 Sequence
C 339	15	28.8	9490	4	AF091049	AF091049 Canis fam
C 340	15	28.8	9926	9	AF257305	AF257305 Homo sapi
C 341	15	28.8	9959	1	AE014682	AE014682 Bifidobac
C 342	15	28.8	10057	1	AE004871	AE004871 Pseudomon
C 343	15	28.8	10151	8	AF484941	AF484941 Fusarium
C 344	15	28.8	10205	1	AE004830	AE004830 Pseudomon
C 345	15	28.8	10210	1	AE014776	AE014776 Bifidobac
C 346	15	28.8	10317	1	AE011691	AE011691 Xanthomon
C 347	15	28.8	10452	1	AE012453	AE012453 Xanthomon
C 348	15	28.8	10510	1	AE012870	AE012870 Chlorobiu
C 349	15	28.8	10510	1	AE000088	AE000088 Rhizobiu
C 350	15	28.8	10734	1	AE012845	AE012845 Myxococcu
C 351	15	28.8	10847	1	AF127082	AF127082 Chlorobiu
C 352	15	28.8	10930	1	AE0012926	AE0012926 Chlorobiu
C 353	15	28.8	11415	1	AE004889	AE004889 Pseudomon
C 354	15	28.8	11431	1	AF372655	AF372655 Rhizobiu
C 355	15	28.8	11455	1	AE001906	AE001906 Deinococc
C 356	15	28.8	11462	1	AE011895	AE011895 Xanthomon
C 357	15	28.8	11684	1	AE005903	AE005903 Caulobact

358	15	28.8	11847	6	AX254620	AX254620 Sequence	C 431	15	28.8	100986	8	AC105363	AC105363 Oryza sat
359	15	28.8	11855	2	AC014860	AC014860 Drosophila	432	15	28.8	103476	9	AC123912	AC123912 Homo sapi
360	15	28.8	12201	1	AE012255	AE012255 Xanthomon	433	15	28.8	103814	2	AC009210	AC009210 Drosophila
361	15	28.8	12277	1	AF340166	AF340166 Streptomy	434	15	28.8	103814	2	AC009210	AC009210 Drosophila
362	15	28.8	12736	1	AE014798	AE014798 Bifidobac	C 435	15	28.8	107661	9	AL162734	AL162734 Human DNA
363	15	28.8	12787	1	AE001720	AE001720 Thermotog	C 436	15	28.8	107947	8	NCB7N14	NCB7N14
364	15	28.8	12789	1	AE012214	AE012214 Xanthomon	C 437	15	28.8	108588	2	AP006098	AP006098 Neurospor
365	15	28.8	14435	2	AC013234	AC013234 Drosophila	C 438	15	28.8	110000	2	AC129386	AC129386 Lotus jap
366	15	28.8	14928	1	AE004676	AE004676 Pseudomon	C 439	15	28.8	110000	2	LMFLCHR15_1	LMFLCHR15_1 of
367	15	28.8	15417	1	AE006966	AE006966 Mycobacte	C 440	15	28.8	110000	2	LMFLCHR16_01	LMFLCHR16_01 of
368	15	28.8	15573	1	RMU66830	U66830 Rhizobium m	441	15	28.8	110000	2	LMFLCHR25_07	LMFLCHR25_07 of
369	15	28.8	16124	6	AX024277	AX024277 Sequence	442	15	28.8	110000	2	LMFLCHR32_13	LMFLCHR32_13 of
370	15	28.8	16623	1	AE005661	AE005661 Escherich	443	15	28.8	110000	2	LMFLCHR32_23	LMFLCHR32_23 of
371	15	28.8	16663	1	AE007074	AE007074 Mycobacte	444	15	28.8	110000	2	LMFLCHR33_24	LMFLCHR33_24 of
372	15	28.8	16929	1	AE006917	AE006917 Mycobacte	445	15	28.8	110000	2	LMFLCHR34_13	LMFLCHR34_13 of
373	15	28.8	18282	1	AF141932	AF141932 Rhizobium	C 446	15	28.8	110000	2	LMFLCHR34_15	LMFLCHR34_15 of
374	15	28.8	23996	2	AC012981	AC012981 Drosophila	447	15	28.8	110000	2	LMFLCHR3_27	LMFLCHR3_27 of
375	15	28.8	24665	9	AL162498	AL162498 Human DNA	448	15	28.8	110000	8	AC145127_14	AC145127_14 of
376	15	28.8	25600	3	UB0442	UB0442 Caenorhabdi	449	15	28.8	110985	8	AP004578	AP004578 Lotus jap
377	15	28.8	29359	3	AC005931	AC005931 Leishmani	450	15	28.8	113459	8	AP006392	AP006392
378	15	28.8	29538	8	APU85909	APU85909 Aureobasidi	C 451	15	28.8	114000	2	AC125412	AC125412
379	15	28.8	30098	2	AC079026	AC079026 Leishmani	C 452	15	28.8	115446	2	AC120402	AC120402
380	15	28.8	31638	1	RMEXPGNS	RMEXPGNS R. meli	C 453	15	28.8	120145	2	AC009842	AC009842
381	15	28.8	31880	6	AR204346	AR204346 Sequence	C 454	15	28.8	121038	8	AC003342	AC003342
382	15	28.8	32765	9	HS363A12	HS363A12 Human DNA	C 455	15	28.8	121160	8	AC005970	AC005970
383	15	28.8	35440	1	AF389112	AF389112 Rhizobium	C 456	15	28.8	121163	9	AC106879	AC106879
384	15	28.8	36526	1	MSCGY42	AD080005 Mycobacte	457	15	28.8	122116	8	AC137634	AC137634
385	15	28.8	36804	1	MTCY210	284395 Mycobacteri	C 458	15	28.8	122390	2	AC135918	AC135918
386	15	28.8	37991	9	AP000565	AP000565 Homo sapi	C 459	15	28.8	123360	9	HS2222513	HS2222513
387	15	28.8	38110	1	MTCY427	Z70692 Mycobacteri	460	15	28.8	124205	8	CNS08CQ	CNS08CQ
388	15	28.8	38653	6	AR285078	AR285078 Sequence	461	15	28.8	124290	8	AP003808	AP003808
389	15	28.8	38721	1	MSGY219	AD000013 Mycobacte	C 462	15	28.8	124711	8	AP003808	AP003808
390	15	28.8	38874	3	LMFLJ377	AL138558 Leishmani	C 463	15	28.8	126631	8	AP003379	AP003379
391	15	28.8	39160	1	MTCY10H4	Z80233 Mycobacteri	C 464	15	28.8	126667	8	AC0066234	AC0066234
392	15	28.8	40013	3	LMFLJ302	AL596287 Leishmani	C 465	15	28.8	126667	8	AC0066234	AC0066234
393	15	28.8	40374	3	LMFLJ302	AL59781 Leishmani	C 466	15	28.8	131768	2	AP005476	AP005476
394	15	28.8	41097	1	AF016585	AF016585 Streptomy	C 467	15	28.8	132575	10	AC117261	AC117261
395	15	28.8	41285	3	AC005930	AC005930 Leishmani	468	15	28.8	132790	9	HS167A14	HS167A14
396	15	28.8	41400	8	AP005254	AP005254 Oryza sat	C 469	15	28.8	133767	8	AP003842	AP003842
397	15	28.8	42498	1	MCLC622	Z95398 Mycobacteri	470	15	28.8	134804	8	AP004888	AP004888
398	15	28.8	43417	2	BX294390	BX294390 Homo sapi	471	15	28.8	135782	2	AC125385	AC125385
399	15	28.8	44366	1	AP003016	AP003016 Mesophilic	C 472	15	28.8	135043	2	AC090974	AC090974
400	15	28.8	44735	5	AF456419	AF456419 Takifluga	C 473	15	28.8	139482	2	AL133333	AL133333
401	15	28.8	55593	1	BSAU2571	AJ002571 Bacillus	C 474	15	28.8	139520	5	AL935199	AL935199
402	15	28.8	59414	2	AC091029	AC091029 Homo sapi	C 475	15	28.8	139645	2	AC105932	AC105932
403	15	28.8	59580	9	AC106875	AC106875 Homo sapi	C 476	15	28.8	139841	2	OSJ000054	OSJ000054
404	15	28.8	63809	2	AC135769	AC135769 Rattus no	477	15	28.8	140583	2	AP005413	AP005413
405	15	28.8	66800	10	AL732607	AL732607 Mouse DNA	C 478	15	28.8	140688	2	AP003801	AP003801
406	15	28.8	68750	1	AF210843	AF210843 Sorangium	479	15	28.8	142133	2	AC135014	AC135014
407	15	28.8	68750	6	AR193029	AR193029 Sequence	C 480	15	28.8	142184	2	AC037428	AC037428
408	15	28.8	68750	6	AR199551	AR199551 Sequence	481	15	28.8	142195	2	AC010669	AC010669
409	15	28.8	68750	6	AR199559	AR199559 Sequence	482	15	28.8	143117	2	AP005872	AP005872
410	15	28.8	68750	6	AR199567	AR199567 Sequence	C 483	15	28.8	143515	8	AP003252	AP003252
411	15	28.8	68750	6	AR201097	AR201097 Sequence	C 484	15	28.8	143604	9	AC093610	AC093610
412	15	28.8	68750	6	AR208671	AR208671 Sequence	C 485	15	28.8	143668	8	AP004268	AP004268
413	15	28.8	71989	6	AR172664	AR172664 Sequence	486	15	28.8	143809	2	AC132485	AC132485
414	15	28.8	72887	2	AC017882	AC017882 Drosophila	487	15	28.8	145003	2	AC130604	AC130604
415	15	28.8	79422	2	AC022750	AC022750 Homo sapi	488	15	28.8	145205	2	AL935049	AL935049
416	15	28.8	80609	1	AF116907	AF116907 Rhodococc	489	15	28.8	145319	2	AP005608	AP005608
417	15	28.8	80610	1	AP001204	AP001204 Rhodococc	C 490	15	28.8	145414	9	HS78P24	HS78P24
418	15	28.8	81777	2	AC144927	AC144927 Mus muscu	491	15	28.8	145662	8	CNS08CAV	CNS08CAV
419	15	28.8	82010	8	AB023042	AB023042 Arabidops	C 492	15	28.8	145668	8	AC105364	AC105364
420	15	28.8	83925	8	NCB23421	NCB23421 Neurospor	493	15	28.8	146576	8	AC074282	AC074282
421	15	28.8	84462	3	AB023038	AB023038 Arabidops	C 494	15	28.8	146936	2	AP005010	AP005010
422	15	28.8	84518	3	AC125735	AC125735 Cont	495	15	28.8	147348	2	AC132489	AC132489
423	15	28.8	85167	2	AC131532	AC131532 Homo sapi	496	15	28.8	147750	8	AP003296	AP003296
424	15	28.8	88277	9	AC015811	AC015811 Homo sapi	C 497	15	28.8	148098	9	AC044819	AC044819
425	15	28.8	92053	2	AC017208	AC017208 Drosophila	C 498	15	28.8	151798	2	CNS07EFL	CNS07EFL
426	15	28.8	93475	5	AL672088	AL672088 Zebrafish	C 499	15	28.8	152075	2	AC139631	AC139631
427	15	28.8	94851	9	AC004923	AC004923 Homo sapi	C 500	15	28.8	152081	2	CNS08C82	CNS08C82
428	15	28.8	95521	9	AC134499	AC134499 Rattus no	501	15	28.8	152090	8	AP004669	AP004669
429	15	28.8	96310	9	AL136444	AL136444 Human DNA	C 502	15	28.8	152274	8	AC079853	AC079853
430	15	28.8	96468	2	AC006902	AC006902 Caenorhab	503	15	28.8	152423	8	AC079853	AC079853

C 504	15	28.8 152736	8	AP003566	AC108317	AC108317	Rattus no
C 505	15	28.8 152898	2	AC134432	AF229199	AF229199	Oryza sat
C 506	15	28.8 152945	2	AC130606	AC092389	AC092389	Oryza sat
C 507	15	28.8 153464	2	AC123391	AC135453	AC135453	Rattus no
C 508	15	28.8 153802	9	AL589949	AF281817	AF281817	Tupaia he
C 509	15	28.8 153926	9	AL160059	AL646058	AL646058	Ralestonia
C 510	15	28.8 154181	8	AC131367	BX324145	BX324145	Danio rer
C 511	15	28.8 154862	2	AC139310	AC090443	AC090443	Mus muscu
C 512	15	28.8 155172	2	AC021110	AL805905	AL805905	Zebrafish
C 513	15	28.8 157920	2	AP003564	BX004990	BX004990	Danio rer
C 514	15	28.8 158146	9	AC005076	AL670720	AL670720	Mouse DNA
C 515	15	28.8 158508	2	AC063334	AC010030	AC010030	Drosophill
C 516	15	28.8 158826	2	AC140156	AC113906	AC113906	Rattus no
C 517	15	28.8 158839	8	AC090882	AC129955	AC129955	Bos tauru
C 518	15	28.8 159391	3	AC069199	AL646078	AL646078	Ralestonia
C 519	15	28.8 160931	3	AC093498	AC139412	AC139412	Rattus no
C 520	15	28.8 160984	2	AC021281	AL646073	AL646073	Ralestonia
C 521	15	28.8 161601	3	AC003781	AC034259	AC034259	Homo sapi
C 522	15	28.8 161720	8	AP003581	AC018039	AC018039	Drosophill
C 523	15	28.8 162719	2	AC021221	299111	299111	Bacillus su
C 524	15	28.8 162721	10	AC079638	AC135223	AC135223	Gallus ga
C 525	15	28.8 163012	2	AC142104	AC135111	AC135111	Mus muscu
C 526	15	28.8 163795	2	AC102466	AC135111	AC135111	Mus muscu
C 527	15	28.8 163797	2	AC102268	BSUB0007	BSUB0007	Homo sapi
C 528	15	28.8 163984	5	AL953872	AF288742	AF288742	Homo sapi
C 529	15	28.8 165900	2	AP003942	AP001922	AP001922	Homo sapi
C 530	15	28.8 165900	2	AC011072	AC092256	AC092256	Mus muscu
C 531	15	28.8 166710	3	BX005141	AC107506	AC107506	Rattus no
C 532	15	28.8 166862	9	CNS01DTS	AC127905	AC127905	Rattus no
C 533	15	28.8 166946	2	AC116781	AC117094	AC117094	Rattus no
C 534	15	28.8 167239	3	AC007668	AC118501	AC118501	Rattus no
C 535	15	28.8 167658	2	AC123942	AC120658	AC120658	Rattus no
C 536	15	28.8 168259	2	AC067842	AL845492	AL845492	Mouse DNA
C 537	15	28.8 168764	1	AP002569	AC110454	AC110454	Rattus no
C 538	15	28.8 168764	8	AC087726	AC127039	AC127039	Rattus no
C 539	15	28.8 169137	2	AC117741	AC141209	AC141209	Rattus no
C 540	15	28.8 169142	2	AC053514	AC141058	AC141058	Homo sapi
C 541	15	28.8 169162	8	AP004267	AC136274	AC136274	Rattus no
C 542	15	28.8 169258	9	AC022013	AC131463	AC131463	Rattus no
C 543	15	28.8 169258	8	AC091787	AC140268	AC140268	Rattus no
C 544	15	28.8 169951	2	AC015654	AC093705	AC093705	Drosophill
C 545	15	28.8 171861	2	AC104824	AC098761	AC098761	Rattus no
C 546	15	28.8 172190	3	AC010668	AC095215	AC095215	Rattus no
C 547	15	28.8 172642	8	AP004258	AC130855	AC130855	Rattus no
C 548	15	28.8 173147	2	BX294117	AC108344	AC108344	Rattus no
C 549	15	28.8 173728	9	AC018507	AC094509	AC094509	Rattus no
C 550	15	28.8 175192	3	AC010576	AC105688	AC105688	Rattus no
C 551	15	28.8 176295	2	AC026176	A93003	A93003	Sequence 2
C 552	15	28.8 177189	2	AC142049	AC017740	AC017740	Drosophill
C 553	15	28.8 178009	2	AC119549	AC095707	AC095707	Rattus no
C 554	15	28.8 178073	1	SC059046	AC094219	AC094219	Rattus no
C 555	15	28.8 178249	2	AP006058	AC106678	AC106678	Rattus no
C 556	15	28.8 178473	10	AL672146	AC129456	AC129456	Rattus no
C 557	15	28.8 178776	2	AC068217	AL589661	AL589661	Chimpanze
C 558	15	28.8 179064	2	AL1353807	AC136816	AC136816	Rattus no
C 559	15	28.8 179439	9	AC013897	AC136706	AC136706	Mouse DNA
C 560	15	28.8 179640	9	AC026462	AC132706	AC132706	Mouse DNA
C 561	15	28.8 181178	3	AC023713	AL928716	AL928716	Zebrafish
C 562	15	28.8 181528	3	AC010562	AC130098	AC130098	Rattus no
C 563	15	28.8 183048	2	AC027439	AC127145	AC127145	Rattus no
C 564	15	28.8 183094	10	AC087903	AC111661	AC111661	Rattus no
C 565	15	28.8 183439	3	AC007809	AP000007	AP000007	Pyrococcus
C 566	15	28.8 184600	2	AC117347	AL589699	AL589699	Mouse DNA
C 567	15	28.8 184940	2	AL672264	AC095671	AC095671	Rattus no
C 568	15	28.8 186487	2	BX248402	AC099416	AC099416	Mus muscu
C 569	15	28.8 186609	8	MP0MTG	AC109119	AC109119	Rattus no
C 570	15	28.8 187332	2	AC101292	AC094000	AC094000	Rattus no
C 571	15	28.8 187386	9	AC095507	AL660343	AL660343	Rhizobium
C 572	15	28.8 187439	2	AC107710	AC125564	AC125564	Rattus no
C 573	15	28.8 187516	2	AL928929	AC113573	AC113573	Rattus no
C 574	15	28.8 189043	8	AP004366	AL660366	AL660366	Drosophill
C 575	15	28.8 189050	1	AL646066	AC003696	AC003696	Rattus no
C 576	15	28.8 189219	2	AC009680	AC127791	AC127791	Rattus no

796	14	26.9	968	9	HSN802426	AL157446 Homo sapi	C 869	14	26.9	1970	6	AX262402	AX262402 Sequence
C 797	14	26.9	970	5	AF027978	AF027978 Gallus ga	C 870	14	26.9	1985	8	AF424570	AF424570 Arabidops
C 798	14	26.9	984	3	GBDMS3A	L24501 Globodera r	C 871	14	26.9	2000	6	AX656614	AX656614 Sequence
C 799	14	26.9	991	3	AF171855	AF171855 Drosophila	C 872	14	26.9	2008	9	HSN805524	HSN805524
C 800	14	26.9	1005	14	HIVU76639	UT6633 HIV-2 clone	C 873	14	26.9	2010	9	AK075401	AK075401 Homo sapi
C 801	14	26.9	1020	8	BT002868	BT002868 Arabidops	C 874	14	26.9	2036	6	AX224806	AX224806 Sequence
C 802	14	26.9	1022	3	PTU66467	U66467 Paramedius	C 875	14	26.9	2045	6	DRONOT01	DRONOT01
C 803	14	26.9	1034	6	AX720346	AX720346 Sequence	C 876	14	26.9	2059	6	AX714630	AX714630 Sequence
C 804	14	26.9	1052	5	AF483320	AF483320 Gourea viC	C 877	14	26.9	2059	9	AK057133	AK057133 Homo sapi
C 805	14	26.9	1057	8	AF059582	AF059582 Verticill	C 878	14	26.9	2063	9	BC048985	BC048985
C 806	14	26.9	1058	8	AF087100	AF087100 Compsopeg	C 879	14	26.9	2098	6	AX154565	AX154565 Sequence
C 807	14	26.9	1080	6	E33745	E33745 Fungus-cont	C 880	14	26.9	2107	10	BC023820	BC023820
C 808	14	26.9	1091	3	AF087104	AF087104 Compsopeg	C 881	14	26.9	2119	9	BC030591	BC030591
C 809	14	26.9	1093	3	AF240865	AF240865 Dendrolas	C 882	14	26.9	2125	6	AX647819	AX647819
C 810	14	26.9	1105	8	AF087102	AF087102 Compsopeg	C 883	14	26.9	2132	9	AF311287	AF311287
C 811	14	26.9	1112	8	AF087106	AF087106 Compsopeg	C 884	14	26.9	2133	10	MM0278575	MM0278575
C 812	14	26.9	1120	5	AB059834	AB059834 Conger my	C 885	14	26.9	2141	3	AC014879	AC014879 Drosophila
C 813	14	26.9	1124	8	AF087107	AF087107 Compsopeg	C 886	14	26.9	2171	6	GB1302036	GB1302036
C 814	14	26.9	1140	14	AY223528	AY223528 Human her	C 887	14	26.9	2176	3	BD160313	BD160313
C 815	14	26.9	1161	6	AX451933	AX451933 Sequence	C 888	14	26.9	2176	9	AK024001	AK024001
C 816	14	26.9	1170	3	AF268321	AF268321 Chlorarac	C 889	14	26.9	2254	3	AY058407	AY058407
C 817	14	26.9	1182	6	BD016692	BD016692 Method fo	C 890	14	26.9	2279	3	AF190556	AF190556
C 818	14	26.9	1191	1	AY142858	AY142858 Hellobaci	C 891	14	26.9	2301	3	AK112563	AK112563
C 819	14	26.9	1228	9	AF124348	AF124348 Homo sapi	C 892	14	26.9	2306	3	SC0512793	SC0512793
C 820	14	26.9	1231	8	ZMA430637	ZMA430637 Zea mays	C 893	14	26.9	2307	9	HSN46470	HSN46470
C 821	14	26.9	1290	8	AB038374	AB038374 Trichospo	C 894	14	26.9	2318	10	RATLHCG01	RATLHCG01
C 822	14	26.9	1311	8	HYH224325	HYH224325 Hordeum v	C 895	14	26.9	2332	8	SCGYP7	SCGYP7
C 823	14	26.9	1323	5	AB021166	AB021166 Gallus ga	C 896	14	26.9	2364	3	TBNNKAI	TBNNKAI
C 824	14	26.9	1323	8	AY179982	AY179982 Zea mays	C 897	14	26.9	2478	5	BC045314	BC045314
C 825	14	26.9	1347	3	AB058667	AB058667 Coproterm	C 898	14	26.9	2483	8	AOP250266	AOP250266
C 826	14	26.9	1347	3	AB058668	AB058668 Coproterm	C 899	14	26.9	2496	1	AB084154	AB084154
C 827	14	26.9	1347	3	AB058669	AB058669 Coproterm	C 900	14	26.9	2509	8	AF085063	AF085063
C 828	14	26.9	1347	3	AB058670	AB058670 Coproterm	C 901	14	26.9	2553	8	AK091508	AK091508
C 829	14	26.9	1347	3	AB058671	AB058671 Coproterm	C 902	14	26.9	2634	6	AK072900	AK072900
C 830	14	26.9	1360	5	AB059657	AB059657 Gallus ga	C 903	14	26.9	2634	6	AX469425	AX469425 Sequence
C 831	14	26.9	1375	8	AB086375	AB086375 Sesbania	C 904	14	26.9	2634	6	BD080289	BD080289
C 832	14	26.9	1410	6	AR052124	AR052124 Sequence	C 905	14	26.9	2638	5	AK026796	AK026796
C 833	14	26.9	1410	6	AR104310	AR104310 Sequence	C 906	14	26.9	2664	5	AF026039	AF026039
C 834	14	26.9	1410	6	AR269200	AR269200 Sequence	C 907	14	26.9	2742	6	AX653038	AX653038
C 835	14	26.9	1410	6	AX079089	AX079089 Sequence	C 908	14	26.9	2769	6	AX151623	AX151623
C 836	14	26.9	1414	5	BC043981	BC043981 Xenopus l	C 909	14	26.9	2799	6	AX356915	AX356915
C 837	14	26.9	1439	5	BC008745	BC008745 Homo sapi	C 910	14	26.9	2799	6	AX356916	AX356916
C 838	14	26.9	1439	8	ZMU066404	ZMU066404 Zea mays	C 911	14	26.9	2799	6	BD107372	BD107372
C 839	14	26.9	1496	6	AX405054	AX405054 Sequence	C 912	14	26.9	2814	6	AX714057	AX714057
C 840	14	26.9	1510	8	YSCCTH1	L41133 SaccharomyC	C 913	14	26.9	2814	9	AK006085	AK006085
C 841	14	26.9	1528	3	AB021864	AB021864 Ciona int	C 914	14	26.9	2814	9	AK056120	AK056120
C 842	14	26.9	1546	3	AB021864	AB021864 Caenorhab	C 915	14	26.9	2816	5	AY179883	AY179883
C 843	14	26.9	1576	1	RHMA11NTRA	M28846 Rhizobium m	C 916	14	26.9	2826	5	BC045321	BC045321
C 844	14	26.9	1608	6	AX154567	AX154567 Sequence	C 917	14	26.9	2857	5	BC043853	BC043853
C 845	14	26.9	1617	6	AX122506	AX122506 Sequence	C 918	14	26.9	2861	6	AR072708	AR072708
C 846	14	26.9	1617	6	BD164623	BD164623 Novel pol	C 919	14	26.9	2861	6	E15701	E15701
C 847	14	26.9	1651	9	AF034838	AF034838 Homo sapi	C 920	14	26.9	2879	8	SCYDL234C	SCYDL234C
C 848	14	26.9	1693	6	AX686748	AX686748 Sequence	C 921	14	26.9	2902	10	RATLHCG	RATLHCG
C 849	14	26.9	1695	5	AF164114	AF164114 Fugu rubr	C 922	14	26.9	2921	3	DMY14171	DMY14171
C 850	14	26.9	1698	6	AX507943	AX507943 Sequence	C 923	14	26.9	2924	8	AF2512973	AF2512973
C 851	14	26.9	1715	3	AY051709	AY051709 Drosophila	C 924	14	26.9	2943	3	REU279073	REU279073
C 852	14	26.9	1716	1	AB009289	AB009289 Streptomy	C 925	14	26.9	2960	1	BD012338	BD012338
C 853	14	26.9	1724	6	BC002421	BC002421 Homo sapi	C 926	14	26.9	3078	6	BD158487	BD158487
C 854	14	26.9	1725	6	AX575578	AX575578 Sequence	C 927	14	26.9	3078	6	AK027770	AK027770
C 855	14	26.9	1740	6	AX066613	AX066613 Sequence	C 928	14	26.9	3123	8	OSTMK	OSTMK
C 856	14	26.9	1746	6	AX506758	AX506758 Sequence	C 929	14	26.9	3150	6	AX082312	AX082312
C 857	14	26.9	1746	8	AY142022	AY142022 Arabidops	C 930	14	26.9	3163	6	AX405058	AX405058
C 858	14	26.9	1764	3	AY069488	AY069488 Drosophila	C 931	14	26.9	3163	10	AY188951	AY188951
C 859	14	26.9	1767	14	HS41ETGA	M17547 Epstein-Bar	C 932	14	26.9	3203	3	AX405050	AX405050
C 860	14	26.9	1810	9	BC0097650	BC0097650 Homo sapi	C 933	14	26.9	3205	6	AX400031	AX400031
C 861	14	26.9	1857	9	BC0098877	BC0098877 Homo sapi	C 934	14	26.9	3231	1	PSEACBC	PSEACBC
C 862	14	26.9	1857	6	AK264389	AK264389 Sequence	C 935	14	26.9	3250	9	AK095391	AK095391
C 863	14	26.9	1903	9	AK092990	AK092990 Homo sapi	C 936	14	26.9	3257	6	AX405056	AX405056
C 864	14	26.9	1913	9	BC028158	BC028158 Homo sapi	C 937	14	26.9	3277	6	AY188947	AY188947
C 865	14	26.9	1922	9	AK091772	AK091772 Homo sapi	C 938	14	26.9	3300	1	AF406639	AF406639
C 866	14	26.9	1959	3	LPO430863	LPO430863 Leishmani	C 939	14	26.9	3376	8	AB086374	AB086374
C 867	14	26.9	1961	6	BD156517	BD156517 Primer fo	C 940	14	26.9	3388	1	CBBL5Q	CBBL5Q
C 868	14	26.9	1961	9	AK001634	AK001634 Homo sapi	C 941	14	26.9				

C 942	1	26.9	3452	9	AY188949	Homo sapi
C 943	14	26.9	3468	5	AF176086	AF176086 Gallus ga
C 944	14	26.9	3500	6	AR275002	AR275002 Sequence
C 945	14	26.9	3500	6	AX654236	AX654236 Sequence
C 946	14	26.9	3559	1	AB072376	AB072376 SpHingomo
C 947	14	26.9	3573	2	AC019447	AC019447 Drosophill
C 948	14	26.9	3645	5	AF514281	AF514281 Oreochrom
C 949	14	26.9	3689	6	AR206497	AR206497 Sequence
C 950	14	26.9	3689	6	AX147254	AX147254 Sequence
C 951	14	26.9	3741	1	AB091437	AB091437 Burkhoidle
C 952	14	26.9	3792	6	AX195972	AX195972 Sequence
C 953	14	26.9	3801	6	AX574106	AX574106 Sequence
C 954	14	26.9	3871	6	BD140451	BD140451 Secretd
C 955	14	26.9	3959	9	AY152396	AY152396 Homo sapi
C 956	14	26.9	3960	8	AF064563	AF064563 Hordeum v
C 957	14	26.9	3962	3	AF145671	AF145671 Drosophill
C 958	14	26.9	3979	6	AR308838	AR308838 Sequence
C 959	14	26.9	4012	8	AY037893	AY037893 Phytophill
C 960	14	26.9	4107	3	AY121618	AY121618 Drosophill
C 961	14	26.9	4180	8	AT0510210	AT0510210 Arabidops
C 962	14	26.9	4283	3	AY051813	AY051813 Drosophill
C 963	14	26.9	4404	9	DB7440	DB7440 Human mRNA
C 964	14	26.9	4412	9	BC015052	BC015052 Homo sapi
C 965	14	26.9	4415	6	AX399923	AX399923 Sequence
C 966	14	26.9	4452	9	AK074112	AK074112 Homo sapi
C 967	14	26.9	4462	1	AF126720	AF126720 Mycobacte
C 968	14	26.9	4503	6	AR072702	AR072702 Sequence
C 969	14	26.9	4503	6	E15695	E15695 Human mRNA
C 970	14	26.9	4503	9	HSU65785	HSU65785 Human 150 k
C 971	14	26.9	4620	9	BC033494	BC033494 Homo sapi
C 972	14	26.9	4775	5	CGLR11	Y08109 G.gallus mR
C 973	14	26.9	4810	6	AR206876	AR206876 Sequence
C 974	14	26.9	4810	6	AR308837	AR308837 Sequence
C 975	14	26.9	4810	6	AX367076	AX367076 Sequence
C 976	14	26.9	5163	6	AX308841	AX308841 Sequence
C 977	14	26.9	5228	8	AF072683	AF072683 Emricell
C 978	14	26.9	5242	9	HSM805174	ALB33879 Homo sapi
C 979	14	26.9	5400	1	AY1HUPZ	L25315 Azotobacter
C 980	14	26.9	5400	14	HS1ULZ	M19120 Herpes simp
C 981	14	26.9	5410	1	AE005715	AE005715 Caulobact
C 982	14	26.9	5451	6	AX078430	AX078430 Sequence
C 983	14	26.9	5471	6	AX305572	AX305572 Sequence
C 984	14	26.9	5471	10	MMNFMG	X05640 Mouse NF-M
C 985	14	26.9	5509	1	BBJ224435	AJ224435 Bifidobac
C 986	14	26.9	5509	6	AR308836	AR308836 Sequence
C 987	14	26.9	5509	6	AR318625	AX318625 Sequence
C 988	14	26.9	5612	6	AX406977	AX406977 Sequence
C 989	14	26.9	5966	1	SAMY18862	Y18862 Streptomyce
C 990	14	26.9	6099	14	PMO271595	AY271595 Poinsettia
C 991	14	26.9	6208	6	AX345097	AX345097 Sequence
C 992	14	26.9	6236	6	BD084814	BD084814 Diagnosis
C 993	14	26.9	6236	9	HSNFM	Y0067 Human gene
C 994	14	26.9	6502	6	AX096213	AX096213 Sequence
C 995	14	26.9	6505	9	AB002346	AB002346 Homo sapi
C 996	14	26.9	6505	9	AF033945	AF033945 Homo sapi
C 997	14	26.9	6642	6	I64799	I64799 Sequence 5
C 998	14	26.9	6715	1	BSCGAL	X9339 B.subtilis
C 999	14	26.9	6746	1	AF318616	AF318616 Homo sapi
C 1000	14	26.9	6799	9	HSM802436	AL157424 Homo sapi

ALIGNMENTS

RESULT 1					
LOCUS	CIU39835	1201 bp	mRNA	linear	PLN 18-JAN-1996
DEFINITION	Coccidioides immitis immunoreactive cell wall protein mRNA,				
	complete cds.				
ACCESSION	U39835				
VERSION	U39835.1	GI:1161373			
KEYWORDS					
SOURCE	Coccidioides posadasii				

ORGANISM	Coccidioides posadasii Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE	Dugger,K.O., Villalreal,K.M., Nguyen,A., Zimmermann,C.R., Law,J.H. and Galgiani,J.N.
AUTHORS	1 (bases 1 to 1201)
TITLE	Cloning and sequence analysis of the cDNA for a protein from Coccidioides immitis with immunogenic potential
JOURNAL	Biochem. Biophys. Res. Commun. 218 (2), 485-489 (1996)
MEDLINE	96144702
PUBMED	8614782
REFERENCE	2 (bases 1 to 1201)
AUTHORS	Dugger,K.O. and Galgiani,J.N.
TITLE	Direct Submission
JOURNAL	Submitted (01-NOV-1995) Kris O. Dugger, Infectious Diseases, University of Arizona, 1501 N. Campbell, Tucson, AZ 85724, USA
FEATURES	Location/Qualifiers
source	1..1201 /organism="Coccidioides posadasii" /mol_type="mRNA" /strain="Silveira" /db_xref="taxon:199306" 152..716 /note="33 kDa apoglycoprotein" /codon_start=1 /product="immunoreactive spherule cell wall protein" /protein_id="AAC49140.1" /db_xref="GI:1161374" translation="MPSVALILVAAGLASAQLDPIPCALNCFVNLGNDGCTRLT DPCHSKSEPLPGQITPCVEACPLVARISYNIIVDDKCAKVPITPDVTAAAE PSTTAETAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPT PIASTPAEPFGASNVRAVSVGIAALLGLAAVL"
CDS	/codon_start=1 /product="immunoreactive spherule cell wall protein" /protein_id="AAC49140.1" /db_xref="GI:1161374"
BASE COUNT	249 a 369 c 275 g 308 t
ORIGIN	
Query Match	100.0%; Score 52; DB #: Length 1201;
Best Local Similarity	100.0%; Pred.No.2.7e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Q#	1 ATGCAGTTCTTCACGCTCATCGCTCTGTCCGCTGCCGCTGGCAGTG 52 Db 152 ATGCAGTTCTTCACGCTCATCGCTCTGTCCGCTGCCGCTGGCAGTG 203
RESULT 2	
LOCUS	CIU32518 1234 bp mRNA linear PLN 04-NOV-1996
DEFINITION	Coccidioides immitis antigen 2 (Ag2) mRNA, complete cds.
ACCSSION	U32518
VERSION	U32518.1 GI:1200179
KEYWORDS	
SOURCE	Coccidioides posadasii
ORGANISM	Coccidioides posadasii Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE	1 (bases 1 to 1234)
AUTHORS	Zhu,Y., Yang,C., Magee,D.M. and Cox,R.A.
TITLE	Molecular cloning and characterization of Coccidioides immitis antigen 2 cDNA
JOURNAL	Infect. Immun. 64 (7), 2695-2699 (1996)
MEDLINE	96294782
PUBMED	8694897
REFERENCE	2 (bases 1 to 1234)
AUTHORS	Cox,R.A.
TITLE	Direct Submission
JOURNAL	Submitted (27-JUL-1995) Texas Center for Infectious Disease, Research Immunology, 2303 S.E. Military Drive, San Antonio, TX 78223, USA
COMMENT	On Feb 24, 1996 this sequence version replaced gi:1184058.
FEATURES	Location/Qualifiers
source	1..1234 /organism="Coccidioides posadasii" /mol_type="mRNA"

VERSION AE004517.1 GI:9946710
KEYWORDS
SOURCE
ORGANISM Pseudomonas aeruginosa PA01
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
AUTHORS Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Iarbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z. and Paulsen,I.T.
TITLE Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
JOURNAL Nature 406 (6799), 959-964 (2000)
MEDLINE 20437337
PUBMED 10984043
REFERENCE
AUTHORS 2 (bases 1 to 10822)
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Iarbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saiter,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 10822)
REFERENCE
AUTHORS Pseudomonas aeruginosa Community Annotation Project (PseudocAP)
CONSORTM Direct Submission
TITLE Submitted (04-FEB-2003) Department of Molecular Biology and
JOURNAL Biochemistry, Simon Fraser University, 8888 University Dr.,
Burnaby, British Columbia V5A 1S6, Canada
COMMENT
This represents the February 3, 2003 version of the continually
updated, reviewed, Pseudomonas aeruginosa PA01 genome annotation,
from PseudocAP (see <http://www.pseudomonas.com> for latest updates
and links to alternate annotations). PseudocAP is coordinated by
Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
E.W. Hancock (University of British Columbia, Canada). We welcome
submission through www.pseudomonas.com of any proposed changes.
'Protein name confidence' is used to rate our confidence of the
accuracy of the protein name.
Class 1: Function experimentally demonstrated in P. aeruginosa.
Class 2: Function of highly similar gene experimentally
demonstrated in another organism (and gene context consistent
in terms of pathways its involved in, if known).
Class 3: Function proposed based on presence of conserved amino
acid motif, structural feature or limited sequence similarity
to an experimentally studied gene.
Class 4: Homologs of previously reported genes of unknown function,
or no similarity to any previously reported sequences.

Location/Qualifiers
1..10822
/organism="Pseudomonas aeruginosa PA01"
/mol_type="genomic DNA"
/strain="PA01"
/db_xref="taxon:208964"
81..506
/locus_tag="PA0817"
81..506
/locus_tag="PA0817"
/locus_tag="PA0817"
/note="Protein name confidence: Class 3 (function proposed
based on presence of conserved amino acid motif,
structural feature or limited sequence similarity to an
experimentally studied gene)"
/codon_start=1

gene
CDS
/transl_table=11
/product="probable ring-cleaving dioxygenase"
/protein_id="AAG04206.1"
/db_xref="GI:9946711"
/translation="MSLMPFLAIPVHDLPAARFYGEVFGSEGRSAEHWDFDFFG
HOLVHHPQDSDHAGSNPVDGHDVDFPVGCVLWAMDMLALERLQDGRFVIE
PIYFKQGVGEQATLPLFDPGQNALPEFSFRMGOLFKA"
complement(525..797)
/locus_tag="PA0818"
complement(525..797)
/locus_tag="PA0818"
/note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04207.1"
/db_xref="GI:9946712"
/translation="MPYVLLILALTPALASSATLDERLGLAHAFKPDLDGTTSAK
PASPPERFKLAPLGSVVRKKCATSSLPDEOSSCALEFTDCAO"
953..1246
/locus_tag="PA0819"
953..1246
/locus_tag="PA0819"
/note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04208.1"
/db_xref="GI:9946713"
/translation="WAPCGKVTVAQSFPIPLRRTSQCNHPCATLYLDCNASE
DLDFARTRLSAIGTLVLDLDETEPGESTCRITGSLSLFEAEVARRRR"
1545..2357
/locus_tag="PA0820"
1545..2357
/locus_tag="PA0820"
/note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04209.1"
/db_xref="GI:9946714"
/translation="MPAVIAENDVSIMDEGCAVYHYPRYSAILKPGVYVYVYKGL
TDKSPADRLSAOPHYRGIARIKGYRPPESKQDFALIESFOFEQAVLAKENSY
LEPIPSKASNYMRSGVREIGVTFDLSLRAELKPLQVABSGRALGCFBSRTGGS
LSTYGVAYERVDPTLRQLQAIHGVACACGDFGKAYGEVAKGFIHVNHTQISEFE
APKAVNPETDLVPLCANCHAVHRRRRTLSVDELKLVGRWCAEADVVV"
complement(2464..3552)
/locus_tag="PA0821"
complement(2464..3552)
/locus_tag="PA0821"
/note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04210.1"
/db_xref="GI:9946715"
/translation="MEQFPDLAELAKOPHLLIAGNLIMKSGPEDYIGAVLCLRGTL
YFKAHPTLVRESLCOCFDEPERLAEPILTWIMREPAQKGLTARDTQPLREMGVA
MEDDDHLSCFYTSKSKSDAGAMFDIYCSKQAKMDHLSVLEFSPIILXOEPOL
DPLQDLPIDPARLREPGCYAGAHAYLSTSDNDNDEPSAPMAARKFGIDVGTCLLAN
TPEKPTRIKTVSWTLTLNERNLRLAGSLDLRLQPLPSHFAFYRGGVVIAQAVP
YIADAEDSRPAPVLLNLNHALKGLIRYETIGSLHGSHDGEIRLVGMADOWLKRLDVE
DSEIPEMCDKLISAEPYLDATNTLPERL"
complement(3569..4015)
/locus_tag="PA0822"

CDS	complement(3569..4015) /locus_tag="PA0822" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AAG04211.1" /db_xref="GI:9946716" /translation="MTKMPPOPILDSDQSGNSPHGWIPOWISKYWDDEDEHPFPKPKGK MIRRDVIVQNPPTQDINKOVEMKPPDPNNRQLEDYALIAIGNKQIVEMKPSDCCQGENQSKVPEQAGMAVLAIGMFWLTGRGSRPMIPAY" complement(4037..4270) /locus_tag="PA0823" complement(4037..4270) /locus_tag="PA0823" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AAG04212.1" /db_xref="GI:9946717" /translation="MSDQPSLRGMSPEGRTPVGTLEPRLLDPQDKKVLCSAVCHCS STPNLSQDGLKQACVQARIGELDEVLOGRSP" complement(4267..4785) /locus_tag="PA0824" complement(4267..4785) /locus_tag="PA0824" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AAG04213.1" /db_xref="GI:9946718" /translation="MSGKPAARVDPPTCPVPGHGSNPIVQSGPDVVEFDGLPAARQCD ASAGSPMISAVSSTVLINGLPATVLTSGIGHAGNVIGSGTULIGDVFPTAPAPRAL PLNRNSVPSGSRFQLIDHETGKPVAGRRVRWSSGMAAFPTTDDQMTSWIERPTAE ILVLDVQRCA" complement(4959..5294) /locus_tag="PA0825" complement(4959..5294) /locus_tag="PA0825" /note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"			
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				
gene				
CDS				

AUTHORS	da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Queglio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J., Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Melandis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spindola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Secubal,J.C. and Kitajima,J.P.		
TITLE	Comparison of the genomes of two Xanthomonas pathogens with differing host specificities		
JOURNAL	Nature 417 (6887), 459-463 (2002)		
MEDLINE	22022145		
PUBMED	12024217		
REFERENCE	2 (bases 1 to 13170)		
AUTHORS	da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Queglio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J., Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Melandis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spindola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Secubal,J.C. and Kitajima,J.P.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-NOV-2001) Departamento de Bioquimica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil		
FEATURES	Location/Qualifiers		
source	1..13170 /organism="Xanthomonas campestris pv. campestris str. ATCC 33913" /mol_type="genomic DNA" /strain="ATCC 33913" /db_xref="ATCC:33913" /db_xref="taxon:190485" /note="pathovar: campestris" 118..1494 /gene="XCC3308" 118..1494 /gene="XCC3308" /note="putative; ORF located using Glimmer/Genemark" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AA042578.1" /db_xref="GI:21114552" /translation="MSLTSSQYALAKDAYDEPPEGTASSRPVDICGTYRLEYVDS PSGYQIYRIORIDNEIIVAHRCETRELRKQDVTYDGGVVAARHNRQAAEAEELTK HALVYAKLKGDSGETTLVAVTGHSLGSLAVQVTAHHSLSKGETRNAGVAVSLDRRIPE GGTDIVINHVAGDAVSAASKRYGVQVQLVATQQLATLEKAGVDYTRSVGLDVNRIVAK KLGDSHRNHNFLPEDANGKPDRSVLEDPKAOQLAKRPAPMIDKYRDVDEVLGSGVTIG VRGNYGIADGDIDALRGTLAPGAGSAAEMAAARQEOORIONDHAPTYVAGKMPYV TSGESEVAPAPAPGCRFEPASASVNPDLVOAHSKIPAGCSAPAAVATVEARAGI LRADLOOSVSHQDNWVIWQOTPEFRVKNVLDSDSVPLQGESIRSQALDTRQSPBVT QGSPARPM" 1499..1858 /gene="XCC3309" 1499..1858		
gene	1499..1858		
CDS	1499..1858		

```

/gene="XCC3309"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/db_xref="GI:21114553"
/translacion="MTERVAPTPAPVTVTAARCPPRATSRLOPLADANLGDVFPEKL
VNDYFNREIQDFGMLMERAMRWSTIALTVTTLWMIGYRIATGOSRESAMATMIK
AGKVA111S1SCSERS"
/complement(1861..2646)
/gene="XCC3310"
/complement(1861..2646)
/gene="XCC3310"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA042580.1"
/db_xref="GI:21114554"
/translacion="MHGANVQORGLPPACHAASGLKVVFNANGAGAFLEKPMKYLRLH
SPGLAVFALMHPVAPARAASWVDAGVTPPGCOLLEAMRATPGASATVAPCNL
WCTEPALGWTDVHGSAAIDIGAKRMLRDEQAHANGAVSGLORAMODPOORASATIA
VLTWSIGTAANTRIHVNRCALVTNAGPDTTACGTEHVLTPANTVLAASGDDRDG
HCAQJGLRLVSKASSIDLVTGRDHATAPRGMITVGINIAMAP"
/complement(2963..3469)
/gene="XCC3311"
/complement(2963..3469)
/gene="XCC3311"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA042581.1"
/db_xref="GI:21114555"
/translacion="MLKSLMTMTGLLSAPVFAAPAPMORAPGTAAIGAAHTLRO
IPEACALRECVTGNAAOPYFSYRSSTCCPRARFVYTAARASVASGMIFNDYIR
VPSACPSQOAVVRVWRKFAAKPOLDGOSRITLDEAKQOAAAGKLPQVPMPTAQO
TMEGRACO"
/3719..5746
/gene="XCC3312"
/3719..5746
/gene="XCC3312"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="H+ translocating pyrophosphate synthase"
/protein_id="AA042582.1"
/db_xref="GI:21114556"
/translacion="MDEHGLMLALGCAVLAIYGMLSARWVVAQPSGNARNOEIAAA
IOEGARAYLNROYLTISAGVLFVLVGLFSLWYTAIFPAIGAVLSGLVIGMNVSV
RANVRTQAARHGISKAMDVARFGAITMLVAGLGLGAGYFALIQMGLPLEONL
HALVGLAFSSLSIFARLGGIFTKGADVADLVGKVBAGIPEDDPRNPAYIANVG
DNVGCAGMAADLFETVAVTVAIATMLGSLTIAEGSPHVAIPLVLCGVSIYASITGA
AFVKRPGSIMGALYKGVISGVYLAATLYLITQALMRENAHGMASLACLAGLGLV
TGLIYVITERTYTGIGYQVHVAASITGHNITAGLSIMKSTALPVIACVCAIHWG
AFHFGGLVGIATAATAMLSMAGMIVALDVGRIITNAGGIAEMAEIPPEVRVNTDPLD
AVNGTITKAVTKGVAIGSALIALVLFDYTNHLOANPEVAFPLSDHTYVIGLIG
GLIPLYFGAMMEAVGARAAGVAVEEROFRIPIGMTTAPROYDRAVMDLTSARIE
EMIVSLPLVVVPIVGLLGPALGGLIGITVIGLPAISIMTGTGGAMDAKAKYIE
DGHFGKSGEAKKAITGDTGDPYKDTAGRAINPLIKINIVALLVPLL"
/complement(6144..6626)
/gene="XCC3313"
/complement(6144..6626)
/gene="XCC3313"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA042583.1"
/db_xref="GI:21114557"

```

```

/translacion="MSRYLRYIVLMSITLAPALAADPPAFVDAVDMFANGECHEA
FVDLEQRDLDRFDNIICGDFCGGEFSFQPLAFRCSVNRSGVRSCTMTFGASEVSV
DPRSGHLRSDSRWCRCTAPLKAGTRLDEMYRVLAVTNPLFEBPLPGCAPPIYNGLICGL
"
/6826..7362
/gene="ppa"
/note="synonym: XCC3314"
/6826..7362
/gene="ppa"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="inorganic pyrophosphatase"
/protein_id="AA042584.1"
/db_xref="GI:21114558"
/translacion="MGLVLTSGKRLPEEINNVIEIPKDESPVKEVDKASCAIPVDR
ILSTPMRYPNCNGYVPTNLGCGDPADVLVPLPVRSVYRCRPVGLRLRSDENG
DEKILAVPIEKIFSGYAHIEDIDVSSHMERITGHFPHYKOLEKGMVKLDGMGGA
EAKRILVESVERVNSDAP"
/7509..8621
/gene="XCC3315"
/7509..8621
/gene="XCC3315"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="response regulator"
/protein_id="AA042585.1"
/db_xref="GI:21114559"
/translacion="WRILFVGDEASLPDLVDYIAELDNDMQAOQVPDGNASIEAAL
SPFDVIVAPVLPDLTAATLGLQIRTLRDPDTRIALIDAODQRTPPAIIIGVAFRL
PMPLEVLVLEVLTLLEELRDLISNPRRAIGRIEXKSPHLYLSLHAALEEDGA
DAADIAKLIGBPATIAAKVLOLCSNAPFGSGRSIDRTAATRLGATVRLDVLASEV
FSVOTLSAERAPAMORALLSSRLAAKULPRTSALGSAALADIGLLPVRBRE
PVEDEDLGHTGACVILGLGLPMPITIEAFAHFRHFORSLRSFWTGAHVATAL
ASGETVDEEYLTQGVINRLPWRQADTLMLGLAA"
/complement(9251..12142)
/gene="btub"
/note="synonym: XCC3316"
/complement(9251..12142)
/gene="btub"

```

Query Match 36.5%; Score 19; DB 1; Length 13170;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	Db	7138	TCTCGTGGTGGCGGCTC	7120
27	TCTCGTGGTGGCGGCTC	45		
1		1		
2		1		
3		1		
4		1		
5		1		
6		1		
7		1		
8		1		
9		1		
10		1		
11		1		
12		1		
13		1		
14		1		
15		1		
16		1		
17		1		
18		1		
19		1		

RESULT 7
AF416330/c
LOCUS AF416330 76093 bp DNA circular BCT 15-MAY-2003
DEFINITION Ruegeria sp. PR1b plasmid pSD20, complete sequence.
ACCESSION AF416330
VERSION AF416330.1 GI:22726324
KEYWORDS
SOURCE
ORGANISM Ruegeria sp. PR1b
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;
Rhodobacteraceae; Ruegeria.
1 (bases 1 to 76093)
Zhong Z., Caspi R., Helinski D., Knaut V., Sykes S., O'Byrne C.,
Shea T.P., Wilkinson J.E., Deloughery C. and Toukdarian A.
Nucleotide sequence based characterizations of two cryptic plasmids
from the marine bacterium Ruegeria isolate PR1b
Plasmid 49 (3), 233-252 (2003)
2 (bases 1 to 76093)
Zhong Z., Toukdarian A. and Helinski D.R.
Direct Submission

JOURNAL Submitted (31-AUG-2001) Department of Biology, University of
California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0322,
USA

FEATURES
source Location/Qualifiers

CDS
1..76093
/organism="Ruegeria sp. Prib"
/mol_type="genomic DNA"
/strain="Prib"
/db_xref="taxon:185588"
/plasmid="pSD20"
complement(75..1079)
/note="replication protein RepA"
/codon_start=1
/transl_table=11
/product="RB100"
/protein_id="AAN05121.1"
/db_xref="GI:22726325"
/translation="MTGISTAAAGYPPREGAGYFLCDIFDAMPKNKDLASMEHPLFSLG
TRPDRIINYNENCAEITVTTPSKGLATIHDKDILIFCVSQMALNAGROVSRILV
TAHDLVACNETSGDAYRRLREAFERLAGRTITNITTTGHEVTSFGILIESMELR
KARGRMVSAVVTLSDMLYRAVLSKSVLTSPDYFLAKPLERLYELARKRCGQDS
WQISVETLLKSGSASPRVRKMRDMTQSPDPYDISELPGDLIRPAQGRRLVES
APPTVSEAREARRLMPGMVDYHALEADWRAVMARTGAPRLKPPDAFLGWVKKRAOD
"

repeat_unit
1179..1193
/note="1teron1"
repeat_unit
1201..1215
/note="1teron2"
repeat_unit
1223..1237
/note="1teron3"
1546..2955
/note="replication protein RepA"
/codon_start=1
/transl_table=11
/product="RB101"
/protein_id="AAN05122.1"
/db_xref="GI:22726326"
/translation="MAKEIHSVAGPALPPYFNIDDDALAELOPTDARFAETAECC
ARGRADLARGINEGSKSLRFTSEITRVLIPVAQAFRRVLKANPDLPOGRSETE
CGAKRTIDEVLRALFAFPAGQSKAKNYLTPRBEGLAKLVAVANFGKGVGTRTAAL
LAMSALDGYKVLVIDLDSQGSMTSIFGKVEDWQTAFLMARHGHLENGRL
DRGAPQPLDLSLADAMETADQVISTHMPNIDLIGAQLNLYAEFOI PVMRFAES
WKLMDALDRLEADQVLDQYDVFIIDTPALGYLTINGLSADILLVPMGASFLEFDS
TGRFFMDHSTFASIEEGENLAALAGREIGEEMDAYRAVITRYDSDAOCGLATLVO
AYMGPTLSRHODYALIGAGGEVNGIYEADYRPFNRTYARNGETDATTAAAKRL
LTGWRRSELALQERRAS"

CDS
3022..4191
/note="replication protein RepB"
/codon_start=1
/transl_table=11
/product="RB102"
/protein_id="AAN05123.1"
/db_xref="GI:22726327"
/translation="MAKKRRLTPAQPTVLSAERAATPSPARGLSKGVSGAATIAQ
VASDTAAALAELEISGLVENARAGIMTEIPLFAIDETVLRDVRVONNEEDMLLMA
SIHSGQOTPIEVALLQPOGSHGLISGMRLTALRLAETPAVTYKALVTP
ASADAYLSMVEENEIRVNLISHYERARIAMRALEGVFPDOKKALLIGVNSPARKS
KISFVTLSALDEALREPTALPEKLGICLAHLLEKDSFGHLLERLRASDPADAVE
EMTVNGMTARNYNLNRLESDDKVPVPRVQPAVTPPOAAPSFPVQMPRSQA
APLVVPASSSDRVMLNLAIPGLRLSFAEEAGTIELSGEAVTADMTLKLWLSRR"
4513..5487
/note="putative galactosyltransferase"
/codon_start=1
/transl_table=11
/product="RB103"
/protein_id="AAN05124.1"
/db_xref="GI:22726328"
/translation="MTRSKLVAVVVTYNRLDKLKYTLARLLESPAEELAQVAVVNNAS
TDGTADMETOHDPRLVLTSETNSGAGFAMGWRHAVETCAPMLVVMDDGRRPAP
GALAFAHLPEDKMDIAAAYFSPSGEICEMNRPSPNPFMLSEFRTALGSGRSGFH
LDPAVEOGCTSIDVTSFVGFISARALARICYPPPSLFIYGDSDIYTLGLTAAGRI
GFEGRVREHDLTSGGGAVOGRRLTLPUMKYTHRLNLLIRLAAGLWMPVLLLI
LPLWLKSRHYGRDGRIFRRLCHAIVRDLRRRTSVSEFTVAKWSGEG"

CDS

5532..7814
/note="putative polysaccharide export protein"
/codon_start=1
/transl_table=11
/product="RB104"
/protein_id="AAN05125.1"
/db_xref="GI:22726329"
/translation="MIAFHONPOCQDORODDIEREDIGRIGRIDPCTEADMD
OGIVPDYLRAGEILVVDHPRGCKQPVODAPSGVPERKQVDDHTTRPDAQSTYH
GKPDRLDHGNVNDRMGVFDGQNPANENAGDANKLDTVEDETROKVDVDDGRVQER
GRERNPGCHVQOAGIKREGQIACVCEOKTGAORGAHDREPOROKDGGORLODG
PDIAFGAPVSRHLSODRADRTOTGASMSDLDGRCGGLRRKRCGGGHEA
AGECVVIRRAGDGLSKMGMTKTANRPGADMDDQDGLSEATASGDDTPAOGKAD
AENRPGCGGCKGGLGPIKAOKLEQVNRBAEATRLHQRLODOPAPVAPATLK
RRHNGILSVIYCVLPLVLYVYILMAIEDOYVSTSGFVARSQATASEALGLAQ
LVGDGTASDSOVLVAFIQSOEIVEVEREVGIRAFSAHPQDMAFLMPDPTLDELIT
WYORILIGISYESGSGLMEVESAVIDPETAQVITRAIISAQDRINALMEQARDAMR
YARADLDVAQVLAKEALTOFRTRTLDPQADIQRMGMNHLQOOLALALVEYD
LLRGLNSDRLKTAETDIVIRERIDIEROTFASDTEGTGLENYPSLISEVERL
TVDQYAOVYRASLTLEVARDEATRSRLATYIKTLPRESSQYPPRPWAGIAGL
LLLWMSIFVLVYTSIRDS"
7827..8486
/note="polysaccharide export protein"
/codon_start=1
/transl_table=11
/product="RB105"
/protein_id="AAN05126.1"
/db_xref="GI:22726330"
/translation="MIREPLTKYFVWDGORTVIDNLTLPSSGSLALGRNGAGK
STLEIIGIORPDLGRVYTDGMSMPVGLASFPKDSGVNRFARIYGVDTDDL
VAFONFSELAFAFSPMRTYSGMSRLTGGASMGIKFDYTLVDEVTAVGDKSFRDK
SRVIFAEKMTKSSALMVHNLNEVRQFCFACGIVLERGATYFDLDEALALHQAAMMO
G"

CDS

complement(8493..10283)
/note="no homolog identified by sequence similarity;
hypothetical protein"
/codon_start=1
/transl_table=11
/product="RB106"
/protein_id="AAN05127.1"
/db_xref="GI:22726331"
/translation="MITLOSFLTPPDEICSEQALYAHKGAAYVSSQSSRYMERGCE
VRLDTYVFLPSLDTMARGCOLDGLCAEIVGGGRGLGVRYHMEGRSTERSFDVLTLT
PEQPAQDLDLHLDGNSGLLYRLRYVMAADGTVEVSGRPATSLADRLPRLTISIT
TFABTEVQRTVERLETFABEATRDQIRVOVVDVAGAGCAPOLESKALATYADPN
LGAGGFARGLLEAEAGASHCLFEMDDASFMENIARTYFLAHDPAITAVAGMI
INTHKMRMESGAMFDCSCRPVFNVDREBEVEVEMLOAPHIRPTLYGCMWFAP
PIKAVTRPFPFVAGDDINFLANFNIRTLNGVSDODDTEKESQTLVLDLRH
LVQHLTIPELRSALGTAKVAIYPTLRSLRHYSTTAVALQAMQDWKMGQPFENI
DMSERRATIKEMSADAEOWPIAQUVLSHRGVALWHSRARRYWGLFTLNGHILDCGG
ASKLVVDINRRALVIRAFGARVSTINTRDCAVTVARDRRFFTLMMQAKLTWOFR
RDPBLSQNSVAGYESRTTKGFWORILGKA"

CDS

complement(11451..12566)
/note="no homolog identified by sequence similarity; ORF"
/codon_start=1
/transl_table=11
/product="RB107"
/protein_id="AAN05128.1"
/db_xref="GI:22726332"
/translation="MAQHKTLIFHLGDKTGSTSIOTAFAOERVOLODASYVPADLA
MNAMARNFSAYHEAKTPGDKAKAEARITAPKRLHQADSPFVISADEPEWAPSVFO
ELYSRYFGIADDELRLISYVPHARLISGYSERKRLGVNADSMETFVERQDRK
LFTFOFRPAARDAGFDAFQURPMI PAOLHGSVVDVLLHAFGLPTKLIGDEANQ
SLCLEDLRLKVQGRHLTVSADVHTTGMDIMKPADMPANPATKRLHSLANDIR
ATYLEDARALDKTFEGKAPLMERELKAIKXAADTPQRLSEPSDHLAEPLSLRLAK
VINGILOEKGVMPAALLHDKRVEDIKNRLQLPSTEGV"
12776..13981
/note="putative UDP-galactopyranose mutase"
/codon_start=1
/transl_table=11
/product="RB108"
/protein_id="AAN05129.1"
/db_xref="GI:22726333"

CDS

complement(11451..12566)
/note="no homolog identified by sequence similarity; ORF"
/codon_start=1
/transl_table=11
/product="RB109"
/protein_id="AAN05130.1"
/db_xref="GI:22726334"

CDS

complement(11451..12566)
/note="no homolog identified by sequence similarity; ORF"
/codon_start=1
/transl_table=11
/product="RB110"
/protein_id="AAN05131.1"
/db_xref="GI:22726335"

CDS
 /translation="MRHRSRIEQAAMAAOKLIMVAGSGAVIGHHLEAGHDITVD
 ARDHIGNCNTRSDDTGVVHAYGPHIHTDDAEVWVNFEPFLPYKXNKVITSL
 DPNQGRGFSLEPNLHTTINQFGLTRPEADHPTKEADSTIDPQFPEQARFV
 GNDLDFEFGKTYIKOMGVHSELSILKRLPVRENDYNEFKHOGKMPENGATSM
 IGLIDHPTITOLSTSEFREMGECDHVGALDGYDVAITLDEPETH
 OGDYOGCAVMYGESEVPYTRITTEKHFSPMHEGHSUYIFRSRLAEDDIPYPIR
 OVEERLADLVKLAETISGVTFVGLGYRLYLDMDVITREALDTRGFEAKTAAGDP
 IPPFEAPL"
 14006..15517
 /note="no homolog identified by sequence similarity;
 hypothetical protein"

CDS
 /codon_start=1
 /transl_table=11
 /product="RB109"
 /protein_id="AA05130.1"
 /db_xref="GI:22726334"
 /translation="WVPCVENTETRVVMSHSSASTARQAVVUGMRSGTSALAG
 VIARLGCALPEQIMPNANDPNPKGFYESLCAVNLNDLLEHADTWRPPLADNDS
 KIGQQLDGTITMLAOEYGDAPFLKDRICILILEYWRLEAOMETIAPFYIHTHRS
 KEVARSLEKRCMPVLRGLLWRYELEAEQTRGPRAFTSYDQLDQWRGALDIQ
 TRCGIDLPEVTEPEICEVEVEGFLADLRHSTALVCDGSLIQWIGCAQVIGRNAAE
 GEDPGRALDQIRKAFDAMPDFGASAGOEALVAPQAGDAERLAREHAAE
 KDELLOKQOATLOAOVMOAKOMEALAEKTRMEALAEKTRMEALAEKTR
 KAOTRLRLMVOEQKAYEVDLSTALSGORRHAMROEELRAEMQRAVARDQAALV
 HDEETERLRRLEROITIDREVLEORERLROSTWRILTAPLRMVNVGLGRA"
 15514..19611
 /note="putative glycosyl transferase"
 /codon_start=1
 /transl_table=11
 /product="RB110"
 /protein_id="AA05131.1"
 /db_xref="GI:22726335"
 /translation="MISGSKGSEMTTHADLSLFOSDWYLEQNPVADHFGVSALEH
 YHFGREGRSPHPLFVAMVYLSQSEPCSEGLEPLTHLFGMKMRSPHVLNPA
 YLRNDDLMQADINPMLYLQSGMOGRSLHPTDRWVLAEDVAEDAGIPLSLYL
 RGMREGRAHPLFASGWAERFQAGADNPMLHYITTCMRMRPRTTFDARWYHDS

Query Match 36.5%; Score 19; DB 1; Length 76093;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 31 GTTCGCGCGCGCTGCCCA 49
 Db 49047 GTTCGCGCGCGCTGCCCA 49029

RESULT 8
 AP005679/c 134963 bp DNA linear HTG 06-SEP-2002
 LOCUS Oryza sativa (japonica cultivar-group) chromosome 9 clone
 DEFINITION OJ1210_A07, *** SEQUENCING IN PROGRESS ***
 ACCESSION AP005679
 VERSION AP005679.1 GI:22775437
 SOURCE HTG: HTGS PHASE2.
 ORGANISM Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1
 Sasaki, T., Matsumoto, T. and Katayose, Y.
 Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, BAC
 clone: OJ1210_A07
 Published Only in Database (2002)
 2 (bases 1 to 134963)
 Sasaki, T., Matsumoto, T. and Katayose, Y.
 Direct Submission
 Submitted (05-SEP-2002) Takuji Sasaki, National Institute of
 Agricultural Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)
 The nucleotide sequence of this BAC clone was generated by

combining Monsanto and RGP-Japan sequencing data.
 NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have been provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1..134963
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="9"
 /clone="OJ1210_A07"
 /c1one="30063 c 30068 g 37798 t 362 others

BASE COUNT 36672 a 30063 c 30068 g 37798 t 362 others

ORIGIN

Query Match 36.5%; Score 19; DB 2; Length 134963;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 19 CTCATCGCTCTCGTGGCTG 37
 Db 76439 CTCATCGCTCTCGTGGCTG 76421

RESULT 9
 AP003250/c 146670 bp DNA linear PLN 29-MAR-2003
 LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
 DEFINITION PAC clone: P0443D08.
 ACCESSION AP003250 BA000010
 VERSION AP003250.3 GI:15408719
 SOURCE
 KEYWORDS
 ORGANISM

Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1

REFERENCE
 AUTHORS Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
 Katayose, Y., Wu, J., Mimura, Y., Cheng, Z., Nagamura, Y.,
 Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Aikawa, K.,
 Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
 Hamada, M., Harada, C., Hishita, S., Honda, M., Ichikawa, Y.,
 Idonuma, A., Iijima, M., Ikeda, M., Ikeno, M., Itoh, S., Itoh, T.,
 Itoh, Y., Itoh, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S.,
 Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Mehara, T.,
 Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashina, M.,
 Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,
 Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T.,
 Shomura, A., Song, J., Takasaki, Y., Teresawa, K., Tsuji, K., Waki, K.,
 Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K.,
 Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y.,
 Yano, M., Jiang, J. and Gojobori, T.
 The genome sequence and structure of rice chromosome 1
 Nature 420 (6913), 312-316 (2002)

TITLE
 JOURNAL MEDLINE
 PUBLISHED 22337376
 12447438

REFERENCE
 2 (bases 1 to 146670)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Direct Submission
 Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
 Agricultural Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)
 On Aug 31, 2001 this sequence version replaced gi:14522861.

COMMENT

Genes were predicted from the integrated results of the following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI Nonredundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DDBJ accession no. and RGP clone ID. A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein. The orientation of the sequence is from T7 to SP6 of the PAC clone. This sequence of P0443D08 clone has an overlap with OSJNBa0083M16 (DDBJ: AP003214) at the position 1 to 624 and with P0480E02 (DDBJ: AP002913) clone at the position 141,467 to 146,670 of 3' end. The sequence of this clone starts at the position 138,088 of OSJNBa0083M16 and ends at the position 5205 of P0480E02. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/Genomeseq.html>.

FEATURES

source

```
1. 146670
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="P0443D08"
2727..3227
/gene="P0443D08.1"
2727..3227
/note="contains ESTs D48568 (S14863), AU089765 (S14863)
similar to 16.9 kDa heat shock protein
unknown protein"
/codon_start=1
/protein_id="BAB64123.1"
/db_xref="GI:15408720"
/translation="MSLVKLEDTLAFDAMNPFISFGTTVAADAMLASDTSAFANTYLE
SRETAAYVFRADLPAGVKEVEVEDEGVLTITGERSVREKGRSHIRSCA
TFEGRFLPDDAVDLVRASMDGMLTITVPKVTIDQPAIAAAPPVAVAPAVEAK
AIEASG"
complement(join(4071..4126,4283..4398,5525..5601))
/gene="P0443D08.2"
complement(join(4071..4126,4283..4398,5525..5601))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB64124.1"
/db_xref="GI:15408721"
/translation="MSLVKLEDTLAFDAMNPFISFGTTVAADAMLASDTSAFANTYLE
SRETAAYVFRADLPAGVKEVEVEDEGVLTITGERSVREKGRSHIRSCA
TFEGRFLPDDAVDLVRASMDGMLTITVPKVTIDQPAIAAAPPVAVAPAVEAK
AIEASG"
complement(join(4071..4126,4283..4398,5525..5601))
/gene="P0443D08.3"
complement(join(4071..4126,4283..4398,5525..5601))
/note="contains EST AU093701 (C63333)"
/codon_start=1
/product="putative 16.9 kDa heat shock protein"
/protein_id="BAB64125.1"
/db_xref="GI:15408722"
/translation="MSLVRSNVFDPFSLDMDPDSVRSVVPATSDNTAAAFANAR
DKETPESHVFKADLPGVKKEVEVEEGVLTISQGRKEKEDKDKMHRVERSSG
QFMRRLPENAKADVQKASMEGVLTITVPKAEVKRPEVKAIEISG"
complement(8660..10035)
/gene="P0443D08.4"
complement(8660..10035)
/gene="P0443D08.4"
```

gene
CDS

gene

CDS

gene
CDS

gene

CDS

gene
CDS

```
/note="pseudogene
similar to non-LTR retroelement reverse transcriptase"
/pseudo
complement(10583..11035)
/gene="P0443D08.5"
complement(10583..11035)
/gene="P0443D08.5"
/note="contains EST AU092880 (C60687)"
/codon_start=1
/product="putative 16.9 kDa heat shock protein"
/protein_id="BAB64126.1"
/db_xref="GI:15408723"
/translation="MSLVRSNVFDPFSLDMDPDSVRSVVPATSDNTAAAFANAR
DKETPESHVFKADLPGVKKEVEVEEGVLTISQGRKEKEDKDKMHRVERSS
QFMRRLPENAKADVQKASMEGVLTITVPKAEVKRPEVKAIEISG"
13625..14077
/gene="P0443D08.6"
13625..14077
/gene="P0443D08.6"
/note="contains EST AU092880 (C60687)"
/codon_start=1
/product="putative 16.9 kDa heat shock protein"
/protein_id="BAB64127.1"
/db_xref="GI:15408724"
/translation="MSLVRSNVFDPFSLDMDPDSVRSVVPATSDNTAAAFANAR
DKETPESHVFKADLPGVKKEVEVEEGVLTISQGRKEKEDKDKMHRVERSS
QFMRRLPENAKADVQKASMEGVLTITVPKAEVKRPEVKAIEISG"
complement(join(16992..17439,18474..18547))
/gene="P0443D08.7"
complement(join(16992..17439,18474..18547))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB64128.1"
/db_xref="GI:15408725"
/translation="MEGCHGQWNTASHFVDEVEREMERPKVFMVLRLLCCRRK
SYRDPVAAAGEEDDEKSSLLTSLRLEELVDDDDCAIDDAVTCNATSCAK
KDGAAPVVLPPGCLHPVWAPRPHGCVTTSSGGGRCOAQVCCRFRFFGGLRRLLM
RRMRPLVLAIPF"
complement(join(19682..20619,20791..21045,21998..22714,
23306..23926,24506..25205))
/gene="P0443D08.8"
complement(join(19682..20619,20791..21045,21998..22714,
23306..23926,24506..25205))
/gene="P0443D08.8"
/note="contains ESTs AU030772 (E60216), AU030771 (E60216)
similar to wall-associated kinase
unknown protein"
/codon_start=1
/protein_id="BAB64129.1"
/db_xref="GI:15408726"
/translation="MSPSFFVAVVSAWSLALMLAAARGAEEGGGCLSGQKCGDLN
ISSPFWIIQQADKPCGPDLOYVCCNNSVATLRSSTDSGFDIINISYGRMLVFD
VHKLARLNNSTGCSI PVFNTFAKLPITFTISPSHNILVFCTAPAPAOQOGLVE
TRCGNTPFALRGFRFEGDYDYKYLEGCRNSVPLVLEPPGKNAASRYVVLVG
GFLITMDLPPTVYSSAANTAADNRPOGCASTYCCGVTISSPAAVPEQATEKCM
LGQVLYCHNTPFLYLYKPRYRQIIDLTFGNNSLVSDDIHKLDPTVFGSVSEYSC
HVRPNTSSKRVGLPFSISTNNLMLFLVSCNAKALVRGDDDLVETRCGNKTFARVGN
YSDGDPAPFYMGCNATVVPVLGTDARSYEOLIRDFGLTWTQETPSASFLLEPAPA
SSSGPGLPTPCCKLTISYPMLEERPSPGSGPPOLKCNATGAYLTITVEAYRVV
DIFPTGNTHVVDENLPLATGCPAPRNISDGIWGPVFIYSEANAEFLRFLSCSLA
AAAPPGHSLPCDDONSVALVSDHHLHEGIGPGCAFVTPVTVQRNGSMACTIAGM
RSGFLIMAYVSDCRKQYSGNCTYSDDLFRACNCPDGMHPDCKEFRSEBACG
SSGGLITVSTIFIAHMKRKRKOTRDLKOLHMSSSMOSYSKDLLEGSPHIFTYEE
LEBATAGFSASRELGDGFGTVYKGLRDGRVAVAKLYKNYRVRVOPFLNEVDILSR
LHQNVILYIGCTSSRSRDLLEYEYIPNGTVADHLGRLPAGERGLTVPMTJAITET
AEALAYLHAEIIRHDKVTNNILLDNPFYKVAIDFGLSRLPLEVTVSVTPQSTPGV
VDVPEIQCYKLTQKSDYSGFVYLIELISCKPAVDMRSRSDIINAMALNRIONHCV
DOLVDEIGYETDSETRKWDVLAELIAFOLCQMDRSPRIKEVLEVLNCTIKGCECA
EKNNKASPKEDSHLKDLSQYSPDSYTHFHFSQSTHSAVSSNGS-
complement(join(27648..27664,28123..28903))
/gene="P0443D08.9"
complement(join(27648..27664,28123..28903))
```

/gene="P0443D08.9"
/note="hypothetical protein"
/codon_start=1
/db_xref="gi:15408727"
/translation="MHPRLCPLASLLLLCHRAAECEPATCGLTRYPFMIGCP
NFNQSQSSPSSALASGCHPAEVEWNGVAVRSQIIVLSDIVSSFFVAHKRVA
DGDGVCRTDFNIISSLSLSPFTISSNRAICFLVSCNGTEPEIDGLVNATISSCSK
PIYAVLGGIYDRDNPPAIKAGNCTVSYLVLPDPSRANLTAGINYSPPQFKGFLVEMQ
KNGFGDCACNGCCCRYYINDSAAFALCISDGLRSTCPGRSF"
complement(join(129901..29938,29998..30738,30790..30937,
30978..31158,31206..31362,31389..31517,31689..31800,
31910..32176))
/gene="P0443D08.10"

Query Match 36.5%; Score 19; DB 8; Length 146670;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 CGCTCTGCTGCTGCTGCGC 42
|||||
Db 68656 CGCTCTGCTGCTGCTGCGC 68638

RESULT 10 156485 bp DNA linear HTG 11-JUL-2002
AP005509 Oryza sativa (japonica cultivar-group) chromosome 8 clone
LOCUS OJ1003 A09. *** SEQUENCING IN PROGRESS ***
DEFINITION AP005509.1 GI:21728111
VERSION HTG; HTGS_PHASE2.
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1
REFERENCE
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
clone:OJ1003 A09
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 156485)
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..156485
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="8"
/clone="OJ1003 A09"

BASE COUNT 43887 a 34663 c 35167 g 42451 t 317 others
ORIGIN

Query Match 36.5%; Score 19; DB 2; Length 156485;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 ACCGCTCATCGCTCTCGT 32
|||||
Db 91526 ACCGCTCATCGCTCTCGT 91508

RESULT 11 157945 bp DNA linear HTG 12-SEP-2002
AP005742 Oryza sativa (japonica cultivar-group) chromosome 9 clone P0489D11,
LOCUS AP005742.1 GI:22830820
DEFINITION *** SEQUENCING IN PROGRESS ***
ACCESSION AP005742
VERSION HTG; HTGS_PHASE2.
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1
REFERENCE
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, PAC
clone:P0489D11
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 157945)
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Direct Submission
JOURNAL Submitted (11-SEP-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..157945
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="9"
/clone="P0489D11"

BASE COUNT 45183 a 35304 c 34626 g 42530 t 302 others
ORIGIN

Query Match 36.5%; Score 19; DB 2; Length 157945;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 CTCATGCTCTGCTGCTG 37
|||||
Db 2999 CTCATGCTCTGCTGCTG 3017

RESULT 12 172081 bp DNA linear HTG 21-FEB-2002
AP004776 Oryza sativa (japonica cultivar-group) chromosome 2 clone P0452F04,
LOCUS AP004776.1 GI:18844996
DEFINITION *** SEQUENCING IN PROGRESS ***
ACCESSION AP004776
VERSION HTG; HTGS_PHASE2.
KEYWORDS

SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehmerioideae; Oryzaceae; Oryza.

REFERENCE
AUTHORS
1 Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
clone: P0452F04
Published Only in Database (2002)

JOURNAL
REFERENCE
AUTHORS
TITLE
2 (bases 1 to 172081)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (20-FEB-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kamondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: csasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 172081
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="P0452F04"
BASE COUNT 49203 a 36993 c 37880 g 47919 t 86 others
ORIGIN

Query Match 36.5%; Score 19; DB 2; Length 172081;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 CGCTCTCGTCGTCGCCGC 42
|||
Db 121561 CGCTCTCGTCGTCGCCGC 121543

RESULT 13
AP005937 298900 bp DNA linear BCT 28-MAR-2003
LOCUS
DEFINITION
Bradyrhizobium japonicum USDA 110 DNA, complete genome, section
3/31
ACCESSION
AP005937 BA000040
VERSION
AP005937.1 GI:27348808
KEYWORDS
ORGANISM
Bradyrhizobium japonicum USDA 110
Bradyrhizobium japonicum USDA 110
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE
AUTHORS
1 Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T.,
Sasamoto, S., Watanabe, A., Ideawa, K., Iriguichi, M., Kawashima, K.,
Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M.
and Tabata, S.
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110
JOURNAL
MEDLINE
PUBMED
DNA Res. 9 (6), 189-197 (2002)
22484998
12597275

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T.,
Sasamoto, S., Watanabe, A., Ideawa, K., Iriguichi, M., Kawashima, K.,
Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M.,
Sasamoto, S., Watanabe, A., Ideawa, K., Iriguichi, M., Kawashima, K.,

TITLE
Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL
MEDLINE
PUBMED
DNA Res. 9 (6), 225-256 (2002)
12597279

REFERENCE
AUTHORS
3 (bases 1 to 298900)
Kaneko, T.
Direct Submission
Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research: 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp,
URL: http://www.kazusa.or.jp/rhizobae/,
Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)
Location/Qualifiers
1. 298900
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
/compilment (114, .329)
/gene="bs10560"
complement (114, .329)
/note="ORF ID: bs10560
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAC45826.1"
/db_xref="GI:27348809"
/translation="MTRFLFVLSLLCSGASAOQPGHDACADVTRFCRPVNNNGD
QAVLACKQNRTRLKSCDKVLTGHGQ"
complement (495, .818)
/gene="b110561"
complement (495, .818)
/note="ORF ID: b110561
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAC45826.1"
/db_xref="GI:27348810"
/translation="MTKSTFKKSLTSSLSKINDTSLALGFAIVSVITSSPAPS
AEAOQCTGDAFLCSSEIPVIFKTIACMKHRSLSGCRVAVMDKLAKASRKVAD
AQQNQ"
1073, .2128
/gene="b110562"
1073, .2128
/gene="b110562"
/note="ORF ID: b110562
putative polysaccharide deacetylase"
/codon_start=1
/transl_table=1
/protein_id="BAC45827.1"
/db_xref="GI:27348811"
/translation="WRNALGLMLASVAVAIVIAAGWFYYSADPGAPKTVAAAPAD
VLAQQTIAKDDVATTAIAAKPTVAOAAPPAVAPVQPAQACNPAALGVARV
VEIDTTGPGFGEHFKEQFDLTDEKLVLFDDQPMVPYPAVVKALADECTKGLFFS
VGKATYHPETILROVLQOHTGTVGTHVSHVNLNSKPTBOOVKDEVEGFAVFPALG
TNAPRPFEPOLQNPAMVYVFCGRVAMFSTIDSDPFRGATPEKIVETVWVRDLK
LGCITLIMHPQKHTGSEAMALALRLKAGGIVQVQAKITFTLPEYDEALLDLQKV
PTSSARPTGSVVQVSO"
complement (2294, .2662)
/gene="b110563"
complement (2294, .2662)
/gene="b110563"
/note="ORF ID: b110563
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAC45828.1"

gene
CDS

/db_xref="GI:27348812"
/translation="MKRIIVAAVVLASASQOAGPRLSLATSEQASAPQATATTA
VTTQVSEAPVTEAPKVDPRPAISAPAPAAVATTSPAAATATKTAKVMPK
KRTWTEGRIISELHRHGIYM"
complement(2830..2905)
/note="crlnt-cgu"
complement(3005..4747)
/gene="b110564"
complement(3005..4747)
/note="ORF_ID:b110564
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC45829.1"
/db_xref="GI:27348813"
/translation="MLRIIVLFLVLTALAAAGAAVADQDGLVLTAGGFRASTLPRF
VELGLFAAAVVLVMSLTTIWRTPGRRLRRSHKRRHARGHAIHGLIAGHGTAL
ARRHAETRRRAHPNDPLALLHAQSAQLGQRDEKORAFRAMAREDRILLGLRGLFI
EQRADDDAVGAMIAEENIKLSPSTASHVLGRCARGDSGLAILDSNLGLI
DKPARRQGVLLTARALELOTMDVARESEVMEIKLAPTLVPAVLAARESEAHQ
VRRAMKLVAAMLANPHDLAEYAVHVKLGDSARQLRVETLAKTPADAGHIEGO
LALARAIDASEFSRARAVLARYNDPTQRYVALMAIEIERTHEGDSGARAMTLRAVR
ARHDPATADGVSDRMRPVSPVGTGLDAFQMTQVASLPSDKGLISSAPEEMILA
APPKRVTPAPNDTLMERPAPASVPAEDONSPPAKKEPKKEPAMTSAEPYTPAPEA
ESSPAATPVFRTRADLCKPAPVPIPAVPIVRAVDPDGIDEGSDEFTQIGTPKA
QAGGKGFWSRWGA"
complement(4784..5995)
/gene="b110565"
complement(4784..5995)
/note="ORF_ID:b110565
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC45830.1"
/db_xref="GI:27348814"
/translation="MADDKPEDAGLAPESGRAPRPPTIDLEATDVSTQPOVAGEPE
APVSEHTEPQAKSEQDAEPERAEQAAPAPASPVSPWYIAPISGVAVAAVI
AVGWMGLPAPVQAPPAAPQVTSATVDALGCVAAVEAAGKPAADPAVAVARIDALEKS
VGLRSDIANLRQSDKTASALNDKASPGASPDLLALYDRIAQLERAGKTERAELA
QOQEKIADARTMDKPLRHVVAAALLDVAVHGDPYEYELAAASLAKPEMLKLOT
PASSGIPTPVALSRELINIVSKLSPDAQSSGGIVIRLQAGASKVIRIERTQGVN
DRGALVAVTAAALRNDVEARELKTLPDADRAPAQWLKARADALAASKRFAD
DAMADLVKPAQ"
complement(6085..6831)
/gene="b110566"
complement(6085..6831)
/note="ORF_ID:b110566
putative utroporphyrinogen III synthase (EC 4.3.1.8)"
/codon_start=1
/transl_table=11
/protein_id="BAC45831.1"
/db_xref="GI:27348815"
/translation="MSILVTRPHPDNEATADNLARGHVVLAPVKEEPAFDESE
ANVSALIVTSANAIRAVAPOLDIGLKLPFPAVCEHTARTARADGPEVVIAGDSGA
ALRPTVRSARDKVLKKSTLLYLAGBLSRDGELTEGFNVVTOITYMAVAVKVL
PRDVCGEFAAHGIEAVLHYSRSASAFUDARDEGVEISALAIPOCCLSFTVASVLRD
AGASQVVAAPKPRDENALFTLERALOTRLA"
6983..8056
/gene="gcp"
6983..8056
/note="ORF_ID:b110567"
/gene="gcp"
/codon_start=1
/transl_table=11
/product="O-sialoglycoprotein endopeptidase"
/protein_id="BAC45832.1"
/db_xref="GI:27348816"
/translation="MLVIGIETCTDETAANIETARPDGIGKILSNIVRSQVDEHAPRG
GVPEIARAHVLDLGGITIDRMRBAGIGFAQLGVAAAGPGLIGGYVGLITTKAKI
ALVHDTPLVAANHLEAHALTRLTLDGIEFPYCLFLASGHTQIIVAVTGQGVYVLGTT

VDDAIGAFDQVAKMLGLPYRGQOVERMAAGGAARFAFRPPROGRPDANFSLGK
TAVNEASRIAEITPDIISDLICASFQAALVETADRLVNGRLFREOFGAPRALVAG
GAANOAIKRGALHDAVARQAKTOLIMPPALCTDNCAMIAMAGLERLAGMTDWEAP
RARMLDANATAPAGYCKTRAG"
8245..9225
/gene="gpsa"
8245..9225
/note="ORF_ID:b110568"
/codon_start=1
/transl_table=11
/product="glycerol-3-phosphate dehydrogenase [NAD(P)+]"
/protein_id="BAC45833.1"
/db_xref="GI:27348817"
/translation="MTAFHSVAVIGAGMGCTALAMVAAAGRSVTLMARNAEHATRIA
STRNPLPGVRLADTIVTSELARADMLLATPQOHARGVNMASHARTPV
VACKGIEHGTHKMTVEVIAEALHAQAPALISGSPADDAVRGLPTATLAKFEELA
SELVALGSPTRFRYHSTDVRCVETIGGAKVLAIAAGIVVGRNIGASALALTTTRGF
SELRILRACGARETSLGSLDILLSTSAOSRNALGIALGRGEAAPAGKLAEG
AFTAPVVELMAARNVMPVSOAVAILDNKLTIDALEGLTPFKAE"
9344..9790
/gene="b110569"
9344..9790
/note="ORF_ID:b110569
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC45834.1"
/db_xref="GI:27348818"
Query Match 36.5%; Score 19; DB 1; Length 298900;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 29 TCGTCGCTGCCGCCCTCGC 47
Db 273395 TCGTCGCTGCCGCCCTCGC 273413
RESULT 14
AY142898
LOCUS AY142898 1500 bp DNA linear BCT 19-DEC-2002
DEFINITION Helicobacillus mobilis Inosine-5'-monophosphate dehydrogenase gene,
partial cds.
ACCESSION AY142898
VERSION AY142898.1 GI:27262441
KEYWORDS
SOURCE Helicobacillus mobilis
ORGANISM Helicobacillus mobilis
REFERENCE 1 (bases 1 to 1500)
AUTHORS Raymond,J., Zhaxybayeva,O., Gogarten,J.P., Gerdes,S.Y. and
Blakenhship,R.E.
TITLE Whole-genome analysis of photosynthetic prokaryotes
JOURNALS Science 298 (5598), 1616-1620 (2002)
MEDLINE 22337798
PUBMED 12446909
REFERENCE 2 (bases 1 to 1500)
AUTHORS Liolios,K.G., Chu,L., Ostrovskaya,O., Mendybaeva,N.,
Koukharenko,V., Gerdes,S., Kyripides,N. and Overbeek,R.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-2002) Bioinformatics, Integrated Genomics, Inc.,
2201 W. Campbell Park Dr., Chicago, IL 60612, USA
FEATURES
source location/Qualifiers
1..1500
/organism="Helicobacillus mobilis"
/mol_type="genomic DNA"
/db_xref="taxon:28064"
/note="RHMO03875"
1..>1500
CDS


```

* 2376 2842: contig of 467 bp in length
* 2843 2922: gap of unknown length
* 2923 3490: contig of 568 bp in length
* 3491 3570: gap of unknown length
* 3571 3985: contig of 415 bp in length
* 3986 4065: gap of unknown length
* 4066 4452: contig of 387 bp in length
* 4453 4532: gap of unknown length
* 4533 5107: contig of 575 bp in length
* 5108 5187: gap of unknown length
* 5188 5875: contig of 688 bp in length
* 5876 5955: gap of unknown length
* 5956 6557: contig of 602 bp in length
* 6558 7066: contig of 429 bp in length
* 7067 7146: gap of unknown length
* 7147 7793: contig of 647 bp in length
* 7794 7873: gap of unknown length
* 7874 8857: contig of 984 bp in length
* 8858 8937: gap of unknown length
* 8938 9508: contig of 571 bp in length
* 9509 9588: gap of unknown length
* 9589 10479: contig of 891 bp in length
* 10480 10559: gap of unknown length
* 10560 12787: contig of 2228 bp in length
* 12788 12867: gap of unknown length
* 12868 14642: contig of 1775 bp in length
* 14643 14722: gap of unknown length
* 14723 16281: contig of 1559 bp in length
* 16282 16361: gap of unknown length
* 16362 17613: contig of 1252 bp in length
* 17614 17693: gap of unknown length
* 17694 19013: contig of 1320 bp in length
* 19014 19094: gap of unknown length
* 19094 21117: contig of 2024 bp in length
* 21118 21197: gap of unknown length
* 21198 22761: contig of 1564 bp in length
* 22762 22841: gap of unknown length
* 22842 25899: contig of 3058 bp in length
* 25900 25979: gap of unknown length
* 25980 31846: contig of 5867 bp in length
* 31847 31926: gap of unknown length
* 31927 37877: contig of 5951 bp in length
* 37878 37957: gap of unknown length
* 37958 44429: contig of 6472 bp in length
* 44430 44509: gap of unknown length
* 44510 51326: contig of 6817 bp in length
* 51327 51406: gap of unknown length
* 51407 57379: contig of 5973 bp in length
* 57380 57459: gap of unknown length
* 57460 58102: contig of 643 bp in length
* 58103 58182: gap of unknown length
* 58183 58489: contig of 307 bp in length
* 58490 58569: gap of unknown length
* 58570 59053: contig of 484 bp in length
* 59054 59133: gap of unknown length
* 59134 59822: contig of 689 bp in length
* 59823 59902: gap of unknown length
* 59903 60285: contig of 383 bp in length
* 60286 60365: gap of unknown length
* 60366 60953: contig of 588 bp in length
* 60954 61033: gap of unknown length
* 61034 61401: contig of 368 bp in length
* 61402 61481: gap of unknown length
* 61482 62093: contig of 612 bp in length
* 62094 62173: gap of unknown length
* 62174 62583: contig of 410 bp in length
* 62584 62663: gap of unknown length
* 62664 62966: contig of 303 bp in length
* 62967 63046: gap of unknown length
* 63047 63353: contig of 307 bp in length
* 63354 63433: gap of unknown length
* 63434 63781: contig of 348 bp in length.

```

```

FEATURES
    source
        Location/Qualifiers
            1..63781
                /organism="Drosophila melanogaster"
                /mol_type="genomic DNA"
                /strain="y; cn bw sp"
                /db_xref="taxon:7227"
                /chromosome="2"
                /map="41C-41C"
                /clone="BACR04F06 (D1057) RPCI-98 04.F.6"
                /clone_1ib="RPCI-98 (Roswell Park Cancer Institute
                Drosophila melanogaster BAC library, partial EcoRI in
                pBAC3.6"
BASE COUNT      17682 a 13292 c 12914 g 16693 t 3200 others
ORIGIN
Query Match      34.6%; Score 18; DB 2; Length 63781;
Best Local Similarity 100.0%; Pred No. 38;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      21 CATCGCTCTGCTGCTGC 38
Db      34873 CATCGCTCTGCTGCTGC 34890

RESULT 17
AC025921
LOCUS      Homo sapiens clone RP11-23F24, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
AC025921
AC025921.1 GI:7259725
VERSION
KEYWORDS
HTG; HTGS PHASE0.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 64271)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE
2 (bases 1 to 64271)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Baerlein,V., Beda,F.,
Boguslavskiy,L., Bouknight,B., Brown,A., Burkett,G.,
Campopiano,A., Casale,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Govette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Iliev,T., Johnson,R., Jones,C., Kann,L., Karatae,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McPeckere,R.,
Meldim,J., Menues,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tasfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viet,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

```

Center project name: L4356
Center Clone name: 23_F_24

* NOTE: This record contains 77 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
1 756: contig of 756 bp in length
757 856: gap of 100 bp
857 1580: contig of 724 bp in length
1581 1680: gap of 100 bp
1681 2388: contig of 708 bp in length
2389 2488: gap of 100 bp
2489 3224: contig of 736 bp in length
3225 3324: gap of 100 bp
3325 4064: contig of 740 bp in length
4065 4164: gap of 100 bp
4165 4905: contig of 741 bp in length
4906 5005: gap of 100 bp
5006 5735: contig of 730 bp in length
5736 5835: gap of 100 bp
5836 6590: contig of 755 bp in length
6591 6690: gap of 100 bp
6691 7402: contig of 712 bp in length
7403 7502: gap of 100 bp
7503 8234: contig of 732 bp in length
8235 8334: gap of 100 bp
8335 9055: contig of 722 bp in length
9056 9156: gap of 100 bp
9157 9906: contig of 750 bp in length
9907 10006: gap of 100 bp
10007 10742: contig of 736 bp in length
10743 10842: gap of 100 bp
10843 11586: contig of 744 bp in length
11587 12413: contig of 727 bp in length
12414 12513: gap of 100 bp
12514 13253: contig of 740 bp in length
13254 13353: gap of 100 bp
13354 14100: contig of 747 bp in length
14101 14200: gap of 100 bp
14201 14943: contig of 743 bp in length
14944 15043: gap of 100 bp
15044 15773: contig of 730 bp in length
15774 15873: gap of 100 bp
15874 16589: contig of 716 bp in length
16590 16689: gap of 100 bp
16690 17418: contig of 729 bp in length
17419 17518: gap of 100 bp
17519 18253: contig of 735 bp in length
18254 18354: gap of 100 bp
18354 19092: contig of 739 bp in length
19093 19192: gap of 100 bp
19193 19923: contig of 731 bp in length
19924 20023: gap of 100 bp
20024 20774: contig of 751 bp in length
20775 20874: gap of 100 bp
20875 21609: contig of 734 bp in length
21609 22462: gap of 100 bp
22462 22562: contig of 754 bp in length
22562 22563: gap of 100 bp
22563 23283: contig of 721 bp in length
23284 23383: gap of 100 bp
23384 24120: contig of 737 bp in length
24121 24220: gap of 100 bp
24221 24972: contig of 752 bp in length

* 24973 25072: gap of 100 bp
25073 25788: contig of 716 bp in length
25789 25888: gap of 100 bp
25889 26634: contig of 746 bp in length
26635 26735: gap of 100 bp
26735 27474: contig of 740 bp in length
27475 27574: gap of 100 bp
27575 28317: contig of 742 bp in length
28317 28416: gap of 100 bp
28417 29131: contig of 715 bp in length
29132 29231: gap of 100 bp
29232 29966: contig of 735 bp in length
29967 30066: gap of 100 bp
30067 30802: contig of 736 bp in length
30803 30902: gap of 100 bp
30903 31633: contig of 731 bp in length
31634 31733: gap of 100 bp
31734 32457: contig of 723 bp in length
32457 32557: gap of 100 bp
32557 33291: contig of 735 bp in length
33292 33391: gap of 100 bp
33392 34117: contig of 726 bp in length
34118 34217: gap of 100 bp
34218 34965: contig of 748 bp in length
34966 35065: gap of 100 bp
35066 35805: contig of 740 bp in length
35806 35905: gap of 100 bp
35906 36649: contig of 744 bp in length
36650 37467: gap of 100 bp
37468 37567: contig of 718 bp in length
37568 38295: gap of 100 bp
38295 38395: contig of 728 bp in length
38396 39123: gap of 100 bp
39124 39223: contig of 728 bp in length
39224 39955: gap of 100 bp
39955 40055: contig of 732 bp in length
40056 40796: gap of 100 bp
40797 40896: contig of 741 bp in length
40897 41629: gap of 100 bp
41630 41729: contig of 733 bp in length
41730 42472: gap of 100 bp
42473 42572: contig of 743 bp in length
42573 43318: gap of 100 bp
43319 43418: contig of 746 bp in length
43419 44166: gap of 100 bp
44167 44266: contig of 748 bp in length
44267 44978: gap of 100 bp
44979 45078: contig of 712 bp in length
45079 45818: gap of 100 bp
45819 45918: contig of 740 bp in length
45919 46636: gap of 100 bp
46637 46737: contig of 718 bp in length
46737 47475: gap of 100 bp
47476 47575: contig of 739 bp in length
47576 48288: gap of 100 bp
48289 48388: contig of 713 bp in length
48389 49105: gap of 100 bp
49106 49205: contig of 717 bp in length
49206 49943: gap of 100 bp
49944 50043: contig of 738 bp in length
50044 50793: gap of 100 bp
50794 50893: contig of 750 bp in length
50894 51657: gap of 100 bp
51658 51757: contig of 764 bp in length
51758 52505: gap of 100 bp
52506 52605: contig of 748 bp in length
52606 53351: gap of 100 bp
53352 53451: contig of 746 bp in length
53452 54214: gap of 100 bp
54215 54314: contig of 763 bp in length
54315 55046: gap of 100 bp
55046 55146: contig of 732 bp in length
55146 55047: gap of 100 bp

```
* 55147 55881: contig of 735 bp in length
* 55882 55981: gap of 100 bp
* 55982 56728: contig of 747 bp in length
* 56729 56828: gap of 100 bp
* 56829 57570: contig of 742 bp in length
* 57571 57670: gap of 100 bp
* 57671 58415: contig of 745 bp in length

Query Match      34.6%; Score 18; DB 2; Length 64271;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      3 GCAGTCTCTCAGCCTCT 20
Db      43891 GCAGTCTCTCAGCCTCT 43908

RESULT 18
AC018480/c 75162 bp DNA linear HTG 21-DEC-1999
LOCUS      Drosophila melanogaster chromosome 2 clone BACR04p15 (D1210)
DEFINITION RPT-98 04.P.15 map 41C-41C strain Y; cn bw sp. *** SEQUENCING IN
ACCESSION  AC018480
VERSION     AC018480.2 GI:6623916
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 75162)
Celisner,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Katta,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacieb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 75162)
Direct Submission
Submitted (13-DEC-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Dec 21, 1999 this sequence version replaced gi:6563452.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 509: contig of 509 bp in length
* 510 589: gap of unknown length
* 590 1119: contig of 530 bp in length
* 1120 1199: gap of unknown length
* 1200 1740: contig of 541 bp in length

* 1741 1820: gap of unknown length
* 1821 2368: contig of 548 bp in length
* 2369 2448: gap of unknown length
* 2449 2955: contig of 507 bp in length
* 2956 3035: gap of unknown length
* 3036 3569: contig of 534 bp in length
* 3570 3649: gap of unknown length
* 3650 4242: contig of 593 bp in length
* 4243 4322: gap of unknown length
* 4323 4709: contig of 387 bp in length
* 4710 4789: gap of unknown length
* 4790 5336: contig of 547 bp in length
* 5337 5416: gap of unknown length
* 5417 5713: contig of 297 bp in length
* 5714 5793: gap of unknown length
* 5794 6355: contig of 562 bp in length
* 6356 6435: gap of unknown length
* 6436 7018: contig of 593 bp in length
* 7019 7099: gap of unknown length
* 7099 7489: contig of 391 bp in length
* 7490 7569: gap of unknown length
* 7570 8012: contig of 443 bp in length
* 8013 8092: gap of unknown length
* 8093 8528: contig of 436 bp in length
* 8529 8608: gap of unknown length
* 8609 9820: contig of 1212 bp in length
* 9821 9900: gap of unknown length
* 9901 10570: contig of 670 bp in length
* 10571 10650: gap of unknown length
* 10651 11391: contig of 741 bp in length
* 11392 11471: gap of unknown length
* 11472 11977: contig of 506 bp in length
* 11978 12057: gap of unknown length
* 12058 12644: contig of 587 bp in length
* 12645 12724: gap of unknown length
* 12725 13189: contig of 465 bp in length
* 13190 13269: gap of unknown length
* 13270 13903: contig of 634 bp in length
* 13904 13983: gap of unknown length
* 13984 14711: contig of 728 bp in length
* 14712 14791: gap of unknown length
* 14792 15507: contig of 716 bp in length
* 15508 15587: gap of unknown length
* 15588 16235: contig of 648 bp in length
* 16236 16315: gap of unknown length
* 16316 17389: contig of 1074 bp in length
* 17390 17469: gap of unknown length
* 17470 18564: contig of 1095 bp in length
* 18565 18644: gap of unknown length
* 18645 20168: contig of 1524 bp in length
* 20169 20248: gap of unknown length
* 20249 20935: contig of 687 bp in length
* 20936 21015: gap of unknown length
* 21016 22156: contig of 1141 bp in length
* 22157 22236: gap of unknown length
* 22237 24521: contig of 2285 bp in length
* 24522 24601: gap of unknown length
* 24602 27047: contig of 2446 bp in length
* 27048 27127: gap of unknown length
* 27128 28250: contig of 1123 bp in length
* 28251 28330: gap of unknown length
* 28331 30388: contig of 2058 bp in length
* 30389 30468: gap of unknown length
* 30469 31598: contig of 1130 bp in length
* 31599 31678: gap of unknown length
* 31679 33076: contig of 1398 bp in length
* 33077 33156: gap of unknown length
* 33157 34906: contig of 1750 bp in length
* 34907 34986: gap of unknown length
* 34987 36702: contig of 1716 bp in length
* 36703 36782: gap of unknown length
* 36783 38644: contig of 1662 bp in length
* 38645 38724: gap of unknown length
```

```

* 38725 41204: contig of 2480 bp in length
* 41205 41284: gap of unknown length
* 41285 44735: contig of 3451 bp in length
* 44736 44815: gap of unknown length
* 44816 45936: contig of 1121 bp in length
* 45937 46016: gap of unknown length
* 46017 47876: contig of 1860 bp in length
* 47877 47956: gap of unknown length
* 47957 50354: contig of 2398 bp in length
* 50355 50434: gap of unknown length
* 50435 54862: contig of 4428 bp in length
* 54863 54942: gap of unknown length
* 54943 58704: contig of 3762 bp in length
* 58705 58784: gap of unknown length
* 58785 62372: contig of 3588 bp in length
* 62373 62452: gap of unknown length
* 62453 67667: contig of 5215 bp in length
* 67668 67747: gap of unknown length
* 67748 68309: contig of 562 bp in length
* 68310 68389: gap of unknown length
* 68390 68995: contig of 606 bp in length
* 68996 69075: gap of unknown length
* 69076 69607: contig of 532 bp in length
* 69608 69687: gap of unknown length
* 69688 69906: contig of 219 bp in length
* 69907 69986: gap of unknown length
* 69987 70544: contig of 558 bp in length
* 70545 70624: gap of unknown length
* 70625 71161: contig of 537 bp in length
* 71162 71241: gap of unknown length
* 71242 71818: contig of 577 bp in length
* 71819 71898: gap of unknown length
* 71899 72368: contig of 470 bp in length
* 72369 72448: gap of unknown length
* 72449 73062: contig of 614 bp in length
* 73063 73142: gap of unknown length
* 73143 73763: contig of 621 bp in length
* 73764 73843: gap of unknown length
* 73844 74543: contig of 700 bp in length
* 74544 74623: gap of unknown length
* 74624 75162: contig of 539 bp in length.

```

FEATURES

SOURCE

```

1. 75162
   location/Qualifiers
   /organism="Drosophila melanogaster"
   /mol_type="genomic DNA"
   /strain="y: cn bw sp"
   /db_xref="taxon:7227"
   /chromosome="2"
   /map="41C-41C"
   /clone="BACR04P15 (D1210) RPCI-98 04.P.15"
   /clone_id="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACe3.6"

```

```

BASE COUNT      19287 a 15828 c 15319 g 19977 t 4751 others
ORIGIN

```

```

Query Match      34.6% Score 18; DB 2; Length 75162;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      21 CATGCTTCGTCGCTGC 38
      |||
Db      65558 CATGCTTCGTCGCTGC 65541

```

```

RESULT 19
AC010568/c
LOCUS
DEFINITION      AC010568 99277 bp DNA linear HTG 29-MAY-2000
Drosophila melanogaster clone RPCI98-1716, *** SEQUENCING IN
PROGRESS ***; 43 unordered pieces.
ACCESSION
AC010568
VERSION
AC010568.5 GI:8101320
KEYWORDS
HTG; HTGS_PHASE1.

```

SOURCE

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 99277)

REFERENCE

AUTHORS

Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K., Bodda, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C., Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C., David, R., Delgado, O., Deshaio, D., Ding, Y., Domah-Rashid, N., Dugan-Rocha, S., Durbin, R.J., Fernandez, C., Ferriguto, D., Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L., Guevara, M., Harris, K., Hernandez, J., Hodgson, A., Hogue, M., Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondajewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, J., Logan, O., Lozano, R.J., Lu, J., Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S., Naeh, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quidley, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, B., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, R., Taber, P., Taylor, T., Vaequez, L., Vinson, R., Vo, O., Wabba, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrenford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.

TITLE

Direct Submission

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 99277)

AUTHORS

Worley, K.C.

JOURNAL

Submitted (16-SEP-1999)

COMMENT

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 29, 2000 this sequence version replaced gi:696776.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: DRFO

Center clone name: RPCI98-1716

Sequencing vector: M13: L08821

Chemistry: Dye-terminator Big Dye 4% of reads

Chemistry: Dye-terminator Big Dye 4% of reads

Assembly program: Phrap; version 0.980611

Consensus quality: 69137 bases at least Q40

Consensus quality: 82549 bases at least Q30

Estimated insert size: 95034; sum-of-contigs estimation

Quality coverage: 2.8x in Q20 bases; sum-of-contigs estimation

NOTE: This is a 'working draft' sequence. It currently consists of 43 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 10767: contig of 10767 bp in length
* 10768 10867: gap of unknown length
* 10868 16333: contig of 5466 bp in length
* 16334 16433: gap of unknown length
* 16434 28333: contig of 11900 bp in length
* 28334 28433: gap of unknown length
* 28434 30659: contig of 2236 bp in length
* 30660 30770: gap of unknown length
* 30771 33570: contig of 2801 bp in length
* 33571 33671: gap of unknown length
* 33672 36538: contig of 2868 bp in length
* 36539 36638: gap of unknown length

```

```

* 36319 38818: contig of 2180 bp in length
* 38819 38918: gap of unknown length
* 40501 40501: contig of 1583 bp in length
* 40602 40601: gap of unknown length
* 42385 42385: contig of 1784 bp in length
* 42485 42485: gap of unknown length
* 45526 45526: contig of 3041 bp in length
* 45527 45626: gap of unknown length
* 45627 47755: contig of 2129 bp in length
* 47756 47855: gap of unknown length
* 47856 49530: contig of 1675 bp in length
* 49531 49630: gap of unknown length
* 49631 51574: contig of 1944 bp in length
* 51575 51674: gap of unknown length
* 53258 53258: contig of 1584 bp in length
* 53359 53358: gap of unknown length
* 55214 55214: contig of 1856 bp in length
* 55315 55314: gap of unknown length
* 55315 58025: contig of 2711 bp in length
* 58025 58125: gap of unknown length
* 58126 60181: contig of 2056 bp in length
* 60182 60281: gap of unknown length
* 60282 62276: contig of 1995 bp in length
* 62277 62376: gap of unknown length
* 62377 63994: contig of 1618 bp in length
* 63995 64094: gap of unknown length
* 64095 66601: contig of 2507 bp in length
* 66602 66701: gap of unknown length
* 66702 67904: contig of 1203 bp in length
* 67905 68004: gap of unknown length
* 68005 69185: contig of 1181 bp in length
* 69186 69285: gap of unknown length
* 69286 70508: contig of 1223 bp in length
* 70509 70608: gap of unknown length
* 70609 72036: contig of 1428 bp in length
* 72037 72136: gap of unknown length
* 72137 74181: contig of 2045 bp in length
* 74182 74281: gap of unknown length
* 74282 75295: contig of 1014 bp in length
* 75296 75395: gap of unknown length
* 75396 76878: contig of 1483 bp in length
* 76879 76978: gap of unknown length
* 76979 78393: contig of 1415 bp in length
* 78394 78493: gap of unknown length
* 80210 80310: contig of 1717 bp in length
* 80311 81493: contig of 1183 bp in length
* 81494 81593: gap of unknown length
* 81594 82783: contig of 1190 bp in length
* 82784 82883: gap of unknown length
* 82884 84809: contig of 1926 bp in length
* 84810 84909: gap of unknown length
* 84910 86592: contig of 1683 bp in length
* 86593 86692: gap of unknown length
* 86693 87752: contig of 1060 bp in length
* 87753 87852: gap of unknown length
* 87853 89155: contig of 1303 bp in length
* 89156 89255: gap of unknown length
* 89256 90443: contig of 1188 bp in length
* 90444 90543: gap of unknown length
* 90544 91893: contig of 1350 bp in length
* 91894 91993: gap of unknown length
* 91994 93070: contig of 1077 bp in length
* 93071 93170: gap of unknown length
* 93171 94175: contig of 1005 bp in length
* 94176 94275: gap of unknown length
* 94276 95483: contig of 1208 bp in length
* 95484 95583: gap of unknown length
* 95584 96799: contig of 1216 bp in length
* 96800 96899: gap of unknown length
* 96900 98099: contig of 1200 bp in length
* 98100 98199: gap of unknown length
* 99277: contig of 1078 bp in length.

```

```

FEATURES
    source
        Location/Qualifiers
            1..99277
                /organism="Drosophila melanogaster"
                /mol_type="genomic DNA"
                /db_xref="taxon:7227"
                /clone="RPC198-1716"
                complement(26106..26205)
                /note="assembly fragment:contig35; vector:Vector_pBACE3.6;
                vector_side:left"
            41196..41295
                /note="assembly fragment:contig36; vector:Vector_pBACE3.6;
                vector_side:right"
    BASE COUNT      28182 a 19521 c 18383 g 28965 t 4226 others
    ORIGIN
        Query Match          34.6%; Score 18; DB 2; Length 99277;
        Best Local Similarity 100.0%; Pred No. 37;
        Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      21 CATCGCTCTCGTCGCTGC 38
Db      35859 CATCGCTCTCGTCGCTGC 35842

RESULT 20
AC008227
LOCUS AC008227
DEFINITION Drosophila melanogaster chromosome 3 clone BACR05008 (DB22) RPC1-98
          05.0.8 map 84p-84d strain y: cn bw sp, *** SEQUENCING IN PROGRESS
          *** 56 unordered pieces.
AC008227
VERSION AC008227.2 GI:5670434
KEYWORDS HTG; HTGS PHASE1.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 101579)
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champagne,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummachi,S.R., Kario,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P.,
Moshirefi,A.R., Moshirefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequerra,A., Sethi,H., Snir,E.,
Svitskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Sequencing of Drosophila melanogaster
2 (bases 1 to 101579)
Unpublished
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champagne,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummachi,S.R., Kario,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P.,
Moshirefi,A.R., Moshirefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequerra,A., Sethi,H., Snir,E.,
Svitskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submisson
Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Aug 2, 1999 this sequence version replaced gi:5649193.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submisson meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```

```

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      589: contig of 589 bp in length
590      669: gap of unknown length
670      1190: contig of 521 bp in length
1191      1270: gap of unknown length
1271      1808: contig of 538 bp in length
1809      1888: gap of unknown length
1889      2483: contig of 595 bp in length
2484      2563: gap of unknown length
2564      3225: contig of 662 bp in length
3226      3305: gap of unknown length
3306      3828: contig of 523 bp in length
3829      3908: gap of unknown length
3909      4435: contig of 527 bp in length
4436      4515: gap of unknown length
4516      5154: contig of 639 bp in length
5155      5234: gap of unknown length
5235      5817: contig of 583 bp in length
5818      5897: gap of unknown length
5898      6983: contig of 1086 bp in length
6984      7063: gap of unknown length
7064      8015: contig of 952 bp in length
8016      8095: gap of unknown length
8096      8784: contig of 689 bp in length
8785      8864: gap of unknown length
8865      9523: contig of 659 bp in length
9524      9603: gap of unknown length
9604      10537: contig of 934 bp in length
10538      10617: gap of unknown length
10618      11429: contig of 812 bp in length
11430      12209: contig of 700 bp in length
12210      12289: gap of unknown length
12290      12938: contig of 649 bp in length
12939      13018: gap of unknown length
13019      14201: contig of 1183 bp in length
14202      14281: gap of unknown length
14282      15114: contig of 833 bp in length
15115      15194: gap of unknown length
15195      16702: contig of 1508 bp in length
16703      16782: gap of unknown length
16783      17607: contig of 825 bp in length
17608      17687: gap of unknown length
17688      19009: contig of 1322 bp in length
19010      19089: gap of unknown length
19090      21081: contig of 1992 bp in length
21082      21161: gap of unknown length
21162      22473: contig of 1312 bp in length
22474      22553: gap of unknown length
22554      23825: contig of 1272 bp in length
23826      23905: gap of unknown length
23906      25283: contig of 1378 bp in length
25284      25363: gap of unknown length
25364      26817: contig of 1454 bp in length
26818      26897: gap of unknown length
26898      28915: contig of 2018 bp in length
28916      28995: gap of unknown length
28996      30314: contig of 1319 bp in length
30315      30394: gap of unknown length
30395      31753: contig of 1359 bp in length
31754      31833: gap of unknown length
31834      33074: contig of 1241 bp in length
33075      33154: gap of unknown length
33155      35326: contig of 2172 bp in length
35327      35406: gap of unknown length
35407      37573: contig of 2167 bp in length
37574      37653: gap of unknown length
37654      39752: contig of 2099 bp in length
39753      39832: gap of unknown length
39833      41517: contig of 1685 bp in length

```

```

*      41518      41597: gap of unknown length
*      41598      44746: contig of 3149 bp in length
*      44747      44826: gap of unknown length
*      44827      48239: contig of 3413 bp in length
*      48240      48319: gap of unknown length
*      48320      52203: contig of 3884 bp in length
*      52204      52283: gap of unknown length
*      52284      59738: contig of 7455 bp in length
*      59739      59818: gap of unknown length
*      59819      67723: contig of 7905 bp in length
*      67724      67803: gap of unknown length
*      67804      72670: contig of 4867 bp in length
*      72671      72750: gap of unknown length
*      72751      91537: contig of 18787 bp in length
*      91538      91617: gap of unknown length
*      91618      92277: contig of 660 bp in length
*      92278      92357: gap of unknown length
*      92358      93027: contig of 670 bp in length
*      93028      93107: gap of unknown length
*      93108      93732: contig of 625 bp in length
*      93733      93812: gap of unknown length
*      93813      94380: contig of 568 bp in length
*      94381      94460: gap of unknown length
*      94461      95199: contig of 739 bp in length
*      95200      95279: gap of unknown length
*      95280      95663: contig of 384 bp in length
*      95664      95743: gap of unknown length
*      95744      96388: contig of 645 bp in length
*      96389      96468: gap of unknown length
*      96469      97219: contig of 751 bp in length
*      97220      97299: gap of unknown length
*      97300      98072: contig of 773 bp in length
*      98073      98152: gap of unknown length
*      98153      98805: contig of 653 bp in length
*      98806      98885: gap of unknown length
*      98886      99435: contig of 550 bp in length
*      99436      99515: gap of unknown length
*      100107      100107: contig of 592 bp in length
*      100187      100187: gap of unknown length
*      100188      100895: contig of 708 bp in length
*      100896      100975: gap of unknown length
*      100976      100976: contig of 604 bp in length.

```

FEATURES

source

```

1. 101579
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3"
/map="84D-84D"
/clone="BACR05008 (D822) RPCI-98 05.O.8"
/clone_id="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBACe3.6"
BASE COUNT      29377 a 20734 c 19813 g 27246 t 4409 others
ORIGIN

```

Query Match

```
Best Local Similarity 34.6%; Score 18; DB 2; Length 101579;
```

```
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      21 CATGCTCTGTCGCTGC 38
Db      67542 CATGCTCTGTCGCTGC 67559

```

RESULT 21

AC098797

LOCUS

DEFINITION

ACCESSION

VERSION

AC098797.3

```

117714 bp DNA linear INV 01-JAN-2003
Leishmania major chromosome 35 clone LB01304 strain Friedlin,
complete sequence.
AC098797
AC098797.3 GI:18139378

```

KEYWORDS HTG
SOURCE Leishmania major
ORGANISM Leishmania major
REFERENCE 1 (bases 1 to 117714)
AUTHORS Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L., McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2001) Seattle Biomedical Research Institution, 4
REFERENCE 2 (bases 1 to 117714)
AUTHORS Myler,P.J., Sisk,E., Nelson,S., Mack,J., Vogt,C., Robertson,L., Worthey,L., Ivens,A., Marty,A., Munden,H., Stuart,K., Rinta,J. and Seyler,A.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-2001) Seattle Biomedical Research Institution, 4
REFERENCE 3 (bases 1 to 117714)
AUTHORS Myler,P.J., Sisk,E., Nelson,S., Mack,J., Vogt,C., Robertson,L., Worthey,L., Ivens,A., Marty,A., Munden,H., Stuart,K., Rinta,J. and Seyler,A.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2002) Seattle Biomedical Research Institution, 4
REFERENCE 4 (bases 1 to 117714)
AUTHORS Myler,P.J., Sisk,E., Nelson,S., Mack,J., Vogt,C., Robertson,L., Worthey,L., Ivens,A., Marty,A., Munden,H., Stuart,K., Rinta,J. and Seyler,A.
TITLE Direct Submission
JOURNAL Submitted (01-JAN-2003) Seattle Biomedical Research Institution, 4
COMMENT Nickerson Street, Seattle, WA 98109-1651, USA
FEATURES
source
1. 117714
/organism="Leishmania major"
/mol_type="genomic DNA"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="35"
/clone="LB01304"
BASE COUNT 22292 a 34596 c 33718 g 27108 t
ORIGIN
Query Match 34.6% Score 18; DB 3; Length 117714;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 TCATCGCTCTCGTCGCTG 37
Db 98830 TCATCGCTCTCGTCGCTG 98847
RESULT 22
LOCUS AC009746
DEFINITION Drosophila melanogaster chromosome 2 clone BACR03D04 (D1041)
AC009746 123010 bp DNA linear HTG 20-MAR-2000
VERSION RPCI-98.03.D.4 map 40D-40D strain Y; cn bw sp; *** SEQUENCING IN
KEYWORDS PROGRESS ***; 80 unordered pieces.
SOURCE AC009746
ORGANISM AC009746.10 GI:7264757
HTG; HTGS PHASE1.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriidae; Drosophilidae; Drosophila.
1 (bases 1 to 123010)
Celiker,S.E., Abayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Burenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequerra,A., Seethi,H., Shtre,E., Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently consists of 80 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

REFERENCE 2 (bases 1 to 123010)
AUTHORS Celiker,S.E., Abayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Burenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequerra,A., Seethi,H., Shtre,E., Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently consists of 80 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 497 577 975 1055 1417 1497 2155 2235 2977 3057 3885 3965 4689 4769 8376 8456 11189 11269 11270 15063 15143 18593 18673 18674 22266 22346 22346 27451 27531 33694 33773 33774 40510 40590 40591 48557 48637 48638 60190 60270 60271 70860 70861 82464 82545 83264 83343 83344

496: contig of 496 bp in length
576: gap of unknown length
974: contig of 398 bp in length
1054: gap of unknown length
1416: contig of 362 bp in length
1496: gap of unknown length
2154: contig of 658 bp in length
2234: gap of unknown length
2976: contig of 742 bp in length
3056: gap of unknown length
3884: contig of 828 bp in length
3964: gap of unknown length
4688: contig of 724 bp in length
4768: gap of unknown length
8375: contig of 3607 bp in length
8455: gap of unknown length
11189: contig of 2734 bp in length
11269: gap of unknown length
15063: contig of 3794 bp in length
15143: gap of unknown length
18593: contig of 3450 bp in length
18673: gap of unknown length
18674: contig of 3552 bp in length
22266: gap of unknown length
22346: contig of 5105 bp in length
27451: gap of unknown length
27531: contig of 6153 bp in length
33694: gap of unknown length
33773: contig of 6737 bp in length
40510: gap of unknown length
40590: gap of unknown length
40591: contig of 7967 bp in length
48557: gap of unknown length
48637: contig of 11553 bp in length
48638: gap of unknown length
60190: gap of unknown length
60270: contig of 10510 bp in length
60271: gap of unknown length
70860: gap of unknown length
70861: contig of 11604 bp in length
82464: gap of unknown length
82545: contig of 719 bp in length
83264: gap of unknown length
83343: contig of 601 bp in length
83344: contig of 601 bp in length


```

* 83945 84024: gap of unknown length
* 84025 84537: contig of 513 bp in length
* 84538 84617: gap of unknown length
* 84618 84853: contig of 236 bp in length
* 84854 84933: gap of unknown length
* 84934 85721: contig of 788 bp in length
* 85722 85801: gap of unknown length
* 85802 86606: contig of 805 bp in length
* 86607 86687: gap of unknown length
* 86687 87380: contig of 694 bp in length
* 87381 87460: gap of unknown length
* 87461 87953: contig of 493 bp in length
* 87954 88033: gap of unknown length
* 88034 88563: contig of 530 bp in length
* 88564 88643: gap of unknown length
* 88644 89377: contig of 734 bp in length
* 89378 89457: gap of unknown length
* 89458 90050: contig of 593 bp in length
* 90051 90130: gap of unknown length
* 90131 90973: contig of 843 bp in length
* 90974 91053: gap of unknown length
* 91054 91730: contig of 677 bp in length
* 91731 91810: gap of unknown length
* 91811 92843: contig of 1033 bp in length
* 92844 92923: gap of unknown length
* 92924 93540: contig of 617 bp in length
* 93541 93621: gap of unknown length
* 93621 94201: contig of 580 bp in length
* 94201 94280: gap of unknown length
* 94281 94644: contig of 364 bp in length
* 94645 94724: gap of unknown length
* 94725 95220: contig of 496 bp in length
* 95221 95300: gap of unknown length
* 95301 95654: contig of 354 bp in length
* 95655 95734: gap of unknown length
* 95735 96325: contig of 591 bp in length
* 96326 96405: gap of unknown length
* 96406 97058: contig of 653 bp in length
* 97059 97138: gap of unknown length
* 97139 97533: contig of 395 bp in length
* 97534 97613: gap of unknown length
* 97614 97970: contig of 357 bp in length
* 97971 98050: gap of unknown length
* 98051 98757: contig of 707 bp in length
* 98758 98837: gap of unknown length
* 98838 99579: contig of 742 bp in length
* 99580 99659: gap of unknown length
* 99660 100277: contig of 618 bp in length
* 100278 100357: gap of unknown length
* 100358 101128: contig of 771 bp in length
* 101129 101208: gap of unknown length
* 101209 102301: contig of 1093 bp in length
* 102302 102381: gap of unknown length
* 102382 102818: contig of 437 bp in length
* 102819 102898: gap of unknown length
* 102899 103305: contig of 407 bp in length
* 103306 103385: gap of unknown length
* 103386 103989: contig of 604 bp in length
* 103990 104069: gap of unknown length
* 104070 104735: contig of 666 bp in length
* 104736 104815: gap of unknown length
* 104816 105416: contig of 601 bp in length
* 105417 105496: gap of unknown length
* 105497 105886: contig of 390 bp in length
* 105887 105966: gap of unknown length
* 105967 106460: contig of 494 bp in length
* 106461 106540: gap of unknown length
* 106541 107486: contig of 946 bp in length
* 107487 107566: gap of unknown length
* 107567 107911: contig of 345 bp in length
* 107912 107991: gap of unknown length
* 107992 108893: contig of 902 bp in length
* 108894 108973: gap of unknown length

```

```

* 108974 109637: contig of 663 bp in length
* 109637 109716: gap of unknown length
* 109717 110025: contig of 309 bp in length
* 110026 110105: gap of unknown length
* 110106 110514: contig of 409 bp in length
* 110515 110594: gap of unknown length
* 110595 110842: contig of 248 bp in length
* 110843 110922: gap of unknown length
* 110923 111715: contig of 793 bp in length
* 111716 111795: gap of unknown length
* 111796 112290: contig of 495 bp in length
* 112291 112370: gap of unknown length
* 112371 113102: contig of 732 bp in length
* 113103 113182: gap of unknown length
* 113183 113446: contig of 264 bp in length
* 113447 113526: gap of unknown length
* 113527 114149: contig of 623 bp in length
* 114150 114229: gap of unknown length
* 114230 114773: contig of 544 bp in length
* 114774 114853: gap of unknown length
* 114854 115250: contig of 397 bp in length
* 115251 115330: gap of unknown length
* 115331 116104: contig of 774 bp in length
* 116105 116184: gap of unknown length
* 116185 116562: contig of 378 bp in length
* 116563 116642: gap of unknown length
* 116643 116857: contig of 215 bp in length
* 116858 116937: gap of unknown length
* 116938 117574: contig of 637 bp in length
* 117575 117654: gap of unknown length
* 117655 118239: contig of 585 bp in length
* 118240 118319: gap of unknown length
* 118320 118629: contig of 310 bp in length
* 118630 118709: gap of unknown length
* 118710 119196: contig of 487 bp in length
* 119197 119276: gap of unknown length

```

Query Match 34.6% Score 18; DB 2; Length 123010;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 CATCGCTCTCGTGCCTGC 38
 Db 30732 CATCGCTCTCGTGCCTGC 30749

RESULT 23
 AC105442/c 124612 bp DNA linear HTG 29-JAN-2002
 LOCUS
 DEFINITION
 AC105442
 Leishmania major chromosome 35 clone LB00329 strain Friedlin, ***
 SEQUENCING IN PROGRESS ***, 3 ordered pieces.

AC105442.2 GI:18390199
 VERSION
 HTG, HTGS PHASE2.
 KEYWORDS
 Leishmania major
 SOURCE
 Leishmania major
 ORGANISM
 Eukaryota; Elenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.

REFERENCE
 1 (bases 1 to 124612)
 Myler,P.J., Sisk,E., Mack,J., Nelson,S., Vogt,C., Robertson,L.,
 Worthey,L., Ivens,A., Marty,A., Munden,H., Stuart,K., Seyler,A. and
 Riney,J.

TITLE
 JOURNAL
 COMMENT
 Submitted (05-JAN-2002) Seattle Biomedical Research Institution, 4
 Nickerson Street, Seattle, WA 98109-1651, USA
 On Jan 29, 2002 this sequence version replaced gi:18071338.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.
 * This sequence will be replaced

* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 103780: contig of 103780 bp in length
* 103781 103930: gap of unknown length
* 103931 106086: contig of 2156 bp in length
* 106087 106236: gap of unknown length
* 106237 124612: contig of 18376 bp in length.
Location/Qualifiers
1. 124612
/organism="Leishmania major"
/mol_type="genomic DNA"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="35"
/clone="LB00329"
BASE COUNT 29211 a 35011 c 36514 g 23576 t 300 others
ORIGIN
Query Match 34.6%; Score 18; DB 2; Length 124612;
Best Local Similarity 100.0%; Pred.No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 20 TCATCGCTCTCGTCGCTG 37
|||||
Db 105161 TCATCGCTCTCGTCGCTG 105144
RESULT 24 127091 bp DNA linear HTG 28-FEB-2000
AC009843/c Drosophila melanogaster chromosome 3 clone BACR01G11 (D1045)
LOCUS RPL1-98 01.G.11 map 80B1-80D2 strain Y; cn bw sp. *** SEQUENCING IN
DEFINITION PROGRESS ***; 93 unordered pieces.
AC009843
AC009843.7 GI:7107697
HTG: HTGS PHASE1.
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 127091)
Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P.,
Moshrefi,M., Nixon,K., Pacleab,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 127091)
Celiner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomocan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleab,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (03-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 28, 2000 this sequence version replaced gi:6980150.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 93 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 562: contig of 562 bp in length
563
643
1276: contig of 634 bp in length
1277
1356: gap of unknown length
1357
1983: contig of 627 bp in length
1984
2064: gap of unknown length
2065
2314: contig of 250 bp in length
2315
2394: gap of unknown length
2395
2865: contig of 472 bp in length
2866
2945: gap of unknown length
2946
3493: contig of 548 bp in length
3494
3573: gap of unknown length
3574
4071: contig of 498 bp in length
4072
4151: gap of unknown length
4152
4897: contig of 746 bp in length
4898
4977: gap of unknown length
4978
5419: contig of 442 bp in length
5420
5489: gap of unknown length
5490
6471: contig of 972 bp in length
6472
6551: gap of unknown length
6552
7505: contig of 954 bp in length
7506
7585: gap of unknown length
7586
8204: contig of 619 bp in length
8205
8284: gap of unknown length
8285
9209: contig of 925 bp in length
9210
9289: gap of unknown length
9290
9811: contig of 692 bp in length
9812
10061: gap of unknown length
10062
11200: contig of 1139 bp in length
11201
11280: gap of unknown length
11281
11987: contig of 707 bp in length
11988
12067: gap of unknown length
12068
12786: contig of 719 bp in length
12787
12867: gap of unknown length
12868
13456: contig of 590 bp in length
13457
13537: gap of unknown length
13538
14333: contig of 797 bp in length
14334
14413: gap of unknown length
14414
15065: contig of 652 bp in length
15066
15145: gap of unknown length
15146
15574: contig of 429 bp in length
15575
15654: gap of unknown length
15655
16170: contig of 516 bp in length
16171
16250: gap of unknown length
16251
16868: contig of 618 bp in length
16869
16948: gap of unknown length
16949
17712: contig of 764 bp in length
17713
17925: gap of unknown length
17926
19276: contig of 1483 bp in length
19277
19355: gap of unknown length
19356
20073: contig of 718 bp in length
20074
20153: gap of unknown length
20154
20797: contig of 644 bp in length
20798
20877: gap of unknown length
20878
22421: contig of 1544 bp in length
22422
22501: gap of unknown length
22502
23477: contig of 976 bp in length
23478
23557: gap of unknown length
23558
24589: contig of 1032 bp in length
24590
24659: gap of unknown length
24660
25381: contig of 712 bp in length
25382
25451: gap of unknown length
25452
26619: contig of 1158 bp in length
26620
26699: gap of unknown length
26700
27921: contig of 1222 bp in length
27922
28001: gap of unknown length
28002
28959: contig of 958 bp in length

```

* 28960 29039: gap of unknown length
* 29040 30564: contig of 1525 bp in length
* 30565 30644: gap of unknown length
* 30645 31644: contig of 1000 bp in length
* 31645 31724: gap of unknown length
* 31725 33109: contig of 1385 bp in length
* 33110 33189: gap of unknown length
* 33190 34898: contig of 1709 bp in length
* 34899 34978: gap of unknown length
* 34979 35797: contig of 819 bp in length
* 35798 35877: gap of unknown length
* 35878 37188: contig of 1311 bp in length
* 37189 37268: gap of unknown length
* 37269 38152: contig of 884 bp in length
* 38153 38232: gap of unknown length
* 38233 40088: contig of 1856 bp in length
* 40089 40168: gap of unknown length
* 40169 41888: contig of 1720 bp in length
* 41889 41968: gap of unknown length
* 41969 43562: contig of 1594 bp in length
* 43563 43642: gap of unknown length
* 43643 44695: contig of 1053 bp in length
* 44696 44775: gap of unknown length
* 44776 46594: contig of 1819 bp in length
* 46595 46674: gap of unknown length
* 46675 48005: contig of 1331 bp in length
* 48006 48085: gap of unknown length
* 48086 49697: contig of 1612 bp in length
* 49698 49777: gap of unknown length
* 49778 50928: contig of 1151 bp in length
* 50929 51008: gap of unknown length
* 51009 53046: contig of 2038 bp in length
* 53047 53126: gap of unknown length
* 53127 54856: contig of 1730 bp in length
* 54857 54936: gap of unknown length
* 54937 56265: contig of 1329 bp in length
* 56266 56345: gap of unknown length
* 56346 57715: contig of 1370 bp in length
* 57716 57795: gap of unknown length
* 57796 59529: contig of 1734 bp in length
* 59530 59609: gap of unknown length
* 59610 61234: contig of 1625 bp in length
* 61235 61314: gap of unknown length
* 61315 63824: contig of 2510 bp in length
* 63825 63904: gap of unknown length
* 63905 64453: contig of 2549 bp in length
* 64454 66533: gap of unknown length
* 66534 71049: contig of 4516 bp in length
* 71050 71129: gap of unknown length
* 71130 75699: contig of 4570 bp in length
* 75700 75779: gap of unknown length
* 75780 79646: contig of 3867 bp in length
* 79647 79726: gap of unknown length
* 79727 84818: contig of 5092 bp in length
* 84819 84898: gap of unknown length
* 84899 90429: contig of 5531 bp in length
* 90430 90509: gap of unknown length
* 90510 93957: contig of 3448 bp in length
* 93958 94037: gap of unknown length
* 94038 100461: contig of 6424 bp in length
* 100462 100541: gap of unknown length
* 100542 106233: contig of 5692 bp in length
* 106234 106313: gap of unknown length
* 106314 106987: contig of 674 bp in length
* 106988 107067: gap of unknown length
* 107068 107780: contig of 713 bp in length
* 107781 107860: gap of unknown length
* 107861 108540: contig of 680 bp in length
* 108541 108620: gap of unknown length
* 108621 109185: contig of 565 bp in length
* 109186 109265: gap of unknown length
* 109266 109901: contig of 636 bp in length
* 109902 109981: gap of unknown length

```

```

Query Match      34.6% Score 18; DB 2; Length 127091;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      21 CATCGCTCTCGCTGCTGC 38
Db      88489 CATCGCTCTCGCTGCTGC 88472

RESULT 25
AC010915
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 137841)
Celisner,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banton,J., Beeson,K.Y., Buesam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dreenek,D., Farfan,D.,
Ferreira,S., Friese,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houch,D., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalili,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phonphanavong,S., Plitman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 4, region 101F-102F
Unpublished (1998)
2 (bases 1 to 137841)
Celisner,S.E., Abayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Buenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomtan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Shit,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (27-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jan 10, 2002 this sequence version replaced gi:7116216.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bddp@fruitfly.berkeley.edu.

```

FEATURES
source
1. 137841
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="4"
/map="101F-102F"
/clone="BACR05L22 (D1068)"
/clone_1ib="RPC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
PBAC3.6)"

BASE COUNT
44191 a 25673 c 24261 g 43716 t
ORIGIN

Query Match 34.6%; Score 18; DB 3; Length 137841;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 21 CATCGCTCTGCTGCTGC 38
|||||
Db 12415 CATCGCTCTGCTGCTGC 12432

RESULT 26
AC134236/c 149349 bp DNA linear PLN 21-DEC-2002
LOCUS
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 3 clone
AC134236
VERSION AC134236.3 GI:27356674
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhacridae; Oryzaceae; Oryza.
1 (bases 1 to 149349)
Wing, R.A., Yu, Y., Soderlund, C., Kim, H.-R., Rambo, T., Currie, J. and
Collura, K.
REFERENCE
AUTHORS
TITLE Rice Genomic Sequence
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 149349)
Wing, R.A., Yu, Y., Soderlund, C., Kim, H.-R., Rambo, T., Sasaki, C.,
Currie, J., and Collura, K.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-2002) Arizona Genomics Institute, University of
AUTHORS Arizona, 303 Forbes, Tucson, AZ 85721, USA
3 (bases 1 to 149349)
Wing, R.A., Yu, Y., Soderlund, C., Kim, H.-R., Rambo, T., Sasaki, C.,
Currie, J., and Collura, K.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2002) Arizona Genomics Institute, University of
AUTHORS Arizona, 303 Forbes, Tucson, AZ 85721, USA
4 (bases 1 to 149349)
Wing, R.A., Yu, Y., Soderlund, C., Kim, H.-R., Rambo, T., Currie, J. and
Collura, K.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2002) Arizona Genomics Institute, University of
AUTHORS Arizona, 303 Forbes, Tucson, AZ 85721, USA
On Dec 21, 2002 this sequence version replaced gi:24270659.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>=30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by more than
one plasmid subclone; and the assembly was confirmed by
restriction digest. There are three areas of single-subclones:
54466-54560, 57568-57653 and 115543-115568. From base 133735-
133753 there are only PCR reads. The assembly overlaps from base
1-22123 with OSJNBa0090010 (accession # AC134237). The overlap is
from 139009-162132 bases on OSJNBa0090010. The assembly overlaps
from base 136321-149349 with OJ1263H11 (accession # AC118980). The

FEATURES
source
1. 149349
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39947"
/chromosome="3"
/clone="OSJNBa0056G13"
/note="(japonica cultivar-group)"

BASE COUNT
39620 a 33572 c 33754 g 42403 t
ORIGIN

Query Match 34.6%; Score 18; DB 8; Length 149349;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 23 TCGCTCTGCTGCTGCGCG 40
|||||
Db 32462 TCGCTCTGCTGCTGCGCG 32445

RESULT 27
AC009353/c 157371 bp DNA linear INV 14-DEC-2001
LOCUS
DEFINITION Drosophila melanogaster, chromosome 2R, region 41E-41E, BAC clone
BACR32F22, complete sequence.
AC009353
VERSION AC009353.8 GI:17737033
KEYWORDS HTG.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 157371)
Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Banson, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Drenek, D., Farfan, D.,
Fiertera, S., Frise, E., Galle, R.F., Gay, S., George, R.A.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
Paclet, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phonannawong, S., Plittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svrtkac, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M., and Venter, J.C.
Sequencing of Drosophila chromosome 2R, region 41E-41E
Unpublished (1998)
2 (bases 1 to 157371)
Celniker, S.E., Adayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummachi, S.R., Katta, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomocan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclet, J.M., Park, S.,
Pfeiffer, B., Poon, L., Segueira, A., Sethi, H., Snir, E.,
Svrtkac, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Dec 14, 2001 this sequence version replaced gi:2264782.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES

Location/Qualifiers
1..157371

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/strain="Y; cn bw sp"

/db_xref="taxon:7227"

/chromosome="2R"

/map="41F-41F"

/clone="BACR32F22 (D910)"

/clone_id="RPCL-98 (Rosewell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

PBAC3.6)"

BASE COUNT

47749 a 29939 c 31348 g 48335 t

ORIGIN

Query Match 34.6%; Score 18; DB 3; Length 157371;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 21 CATGCGCTCGTCTGCG 38

Db 40255 CATGCGCTCGTCTGCG 40238

RESULT 28

AP003935

LOCUS AP003935 159581 bp DNA linear HTG 21-MAR-2002
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 6 clone P0661G04,
*** SEQUENCING IN PROGRESS ***.

ACCESSION AP003935

VERSION AP003935.1 GI:14915718

KEYWORDS HTG; HTGS PHASE2.

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Eurharitoidae; Oryzaceae; Oryza.

REFERENCE

AUTHORS

1 Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa nipponbare (Ga3) genomic DNA, chromosome 6, PAC

clone: P0661G04

2 Published Only in Database (2001)

2 (bases 1 to 159581)

Sasaki, T., Matsumoto, T. and Yamamoto, K.

Direct Submission

Submitted (18-JUL-2001) Takuji Sasaki, National Institute of

Agrobiological Resources, Rice Genome Research Program; Kamondai

2-1-2, Tsukuba, Ibaraki 305-8602, Japan

(E-mail: tsasaki@nias.affrc.go.jp, URL: <http://rpg.dna.affrc.go.jp/>,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)

COMMENT

NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

Location/Qualifiers

1..159581

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosome="6"

/clone="P0661G04"

BASE COUNT

43622 a 35810 c 34889 g 44410 t 850 others

ORIGIN

Query Match 34.6%; Score 18; DB 2; Length 159581;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 17 CTCATGCGCTCTGCTGCG 34

Db 56020 CTCATGCGCTCTGCTGCG 56037

RESULT 29

AC119676

LOCUS AC119676 160573 bp DNA linear PRI 31-JUL-2002

DEFINITION Homo sapiens chromosome 1 clone RP5-820016, complete sequence.

AC119676 AL157905

AC119676.2 GI:22024581

HTG.

KEYWORDS Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 160573)

Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,

Saenphimachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.

and Haugen, E.D.

Direct Submission

Unpublished

2 (bases 1 to 160573)

Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.

Direct Submission

Submitted (30-APR-2002) Genome Center, University of Washington,

Box 352145, Seattle, WA 98195, USA

3 (bases 1 to 160573)

Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,

Saenphimachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.

and Haugen, E.D.

Direct Submission

Submitted (31-JUL-2002) Genome Center, University of Washington,

Box 352145, Seattle, WA 98195, USA

On Jul 31, 2002 this sequence version replaced gi:20340501.

----- Genome Center

Center: University of Washington Genome Center

Center Code: UWGC

Web site: <http://www.genome.washington.edu>

Contact: uwgchgs@u.washington.edu

Drafting Center: SC

----- Project Information

Center Project name: chr-1

Center clone name: RP5-820016 (sc0838)

----- Summary Statistics

Sequencing vector: plasmid; 6% of reads

Chemistry: Dye-terminator Big Dye; 31% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 160468 bases at least Q40

Consensus quality: 160573 bases at least Q30

Insert size: 160573; sum-of-contigs

Quality coverage: 8.7x in Q20 bases; sum-of-contigs

Overlapping Sequences:

5': RP11-486810 AL445933, 2000-bp overlap

3': RP11-216M14 AL451006, 2370-bp overlap

Sequence Quality Assessment:

This entry has been annotated with sequence quality

estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than

1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the

Genbank flat file format but are available as part

of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

HindIII

ECORI

BglIII

SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt

4651	4729	3216	3280	8871	8725
449	<800	2184	2124	5671	5578
512	<800	8065	7987	7274	7290
2814	2828	486	<800	5062	5218
1247	1223	106	<800	3738	3629
4695	4729	2252	2235	4303	4233
3373	3434	3261	3280	2976	2947
3480	3434	1176	1161	3575	3629
1040	1047	485	<800	4090	4059
4205	4315	730	760	4541	4505
1363	1327	1441	1415	5061	5015
2834	2828	8427	8263	1543	1500
580	<800	2059	2019	2059	2069
960	937	488	<800	1452	1438
345	<800	4826	4743	2922	2947
2226	2155	1007	996	793	<800
3655	3625	2022	2019	8159	7991
4826	4729	11962	11922	6687	6701
3185	3237	41	<800	3091	3122
1881	1846	121	<800	571	<800
1499	1482	318	<800	5239	5015
3571	3625	3966	3950	1981	1963
36	<800	6804	6606	272	<800
830	851	130	<800	2227	2226

1257	1223	6623	6606	1692	1722
769	781	7207	7242	11429	11239
1232	1223	6036	5928	15433	15722
3393	3434	2941	2941	5568	5578
2494	2506	9156	9188	8270	8296
4816	4729	2372	2367	11452	11239
367	<800	1292	1256	1221	1219
1107	1653	3253	3280	4183	4233
22	<800	770	760	453	<800
7538	7498	939	996	2058	2069
46	<800	1787	1737	2480	2497
14217	13799	1258	1256	9539	9559
6994	7061	6276	6300	1029	1028
2797	2828	291	<800	4901	5015
2801	2828	9501	9445	4717	4699
1858	1846	1001	996		
6740	6779	7642	7642		
476	<800	6018	5928		
11525	11253	6277	6300		
601	<800	6645	6606		
1236	1223	2105	2124		
4323	4315	3271	3280		
1630	1653	823	828		
1786	1846	528	<800		
2140	2155	1730	1737		
491	<800	7201	7242		
19364	19745	3020	3076		
14536	14543	3533	3528		
2699	2828	1514	1472		
2120	2155				
351	<800				

FEATURES
source Location/Qualifiers
1..160573
/organism="Homo sapiens"

Query Match 34.6% Score 18; DB 9; Length 160573;
Best Local Similarity 100.0%; Pred.No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GCAGTTCCTCAGCCTCT 20
|||||

Db 143869 GCAGTTCCTCAGCTCT 143886

RESULT 30
AC012389
LOCUS
DEFINITION
AC012389 161518 bp DNA linear HTG 24-FEB-2000
Drosophila melanogaster chromosome 3 clone BACR28N05 (D1183)
RFLP: 98.28 N.5 map 100F1-100F5 strain Y; cn bw sp, *** SEQUENCING
IN PROGRESS *** 70 unordered pieces.

ACCESSION
AC012389
KEYWORDS
SOURCE
ORGANISM
HTG: HTGS_PHASE1.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
1 (bases 1 to 161518)
Celisner, S.E., Agdayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,
Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R.,
Moshrefi, M., Nixon, K., Paclebd, J.M., Park, S., Pfeiffer, B.,
Richard, S., Sechi, H., Svitskas, R.R., Wan, K.H., Webster, D.,
Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
Sequencing of Drosophila melanogaster
Unpublished

TITLE
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 161518)
Celisner, S.E., Agdayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomocan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Paclebd, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sechi, H., Sutr, E.,
Svitskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submision
Submitted (27-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 24, 2000 this sequence version replaced gi:6996991.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

TITLE
JOURNAL
COMMENT
1
569: contig of 569 bp in length
570
649: gap of unknown length
650
928: contig of 279 bp in length
929
1008: gap of unknown length
930
1009
1403: contig of 395 bp in length
1404
1483: gap of unknown length
1484
2103: contig of 620 bp in length
2183
2184
2781: gap of unknown length
2782
2861: gap of unknown length
2862
3522: contig of 661 bp in length
3523
3602: gap of unknown length
3603
4192: contig of 590 bp in length
4193
4272: gap of unknown length
4273
4898: contig of 626 bp in length
4899
4978: gap of unknown length
4979
5329: contig of 351 bp in length
5330
5409: gap of unknown length

5410
5954
6034
6948
7029
7029
8874
8954
10162
10242
10242
12996
12996
13076
13076
14726
14805
14805
16263
16263
16342
16342
16474
16474
18475
18475
20725
20725
20805
20805
24209
24209
24289
24289
26485
26485
29109
29109
32551
32551
32630
32630
35639
35639
35719
35719
38937
38937
39017
39017
42731
42731
42811
42811
47072
47072
47151
47151
51485
51485
51565
51565
55313
55313
55393
55393
59361
59361
59441
59441
67280
67280
67359
67359
71455
71455
71535
71535
77602
77602
77682
77682
84384
84384
84463
84463
92634
92634
92713
92713
100750
100750
100751
100751
100831
100831
112677
112677
112757
112757
112758
112758
121915
121915
121995
121995
139644
139644
139723
139723
140359
140359
140439
140439
140440
140440
141138
141138
141219
141219
141882
141882
142062
142062
143327
143327
143407
143407
144194
144194
144328
144328
144394
144394
144775
144775
145021
145021
145100
145100
145588
145588
145668
145668
145669
145669

5953: contig of 544 bp in length
6033: gap of unknown length
6948: contig of 915 bp in length
7028: gap of unknown length
8873: contig of 1845 bp in length
8953: gap of unknown length
10161: contig of 1208 bp in length
10241: gap of unknown length
12996: contig of 2755 bp in length
13076: gap of unknown length
14725: contig of 1649 bp in length
14805: gap of unknown length
16262: contig of 1457 bp in length
16342: gap of unknown length
18474: contig of 2132 bp in length
18554: gap of unknown length
20724: contig of 2170 bp in length
20804: gap of unknown length
24208: contig of 3404 bp in length
24288: gap of unknown length
26484: contig of 2116 bp in length
26484: gap of unknown length
29108: contig of 2624 bp in length
29188: gap of unknown length
32550: contig of 3362 bp in length
32630: gap of unknown length
35639: contig of 3006 bp in length
35719: gap of unknown length
38937: contig of 3218 bp in length
39017: gap of unknown length
42730: contig of 3713 bp in length
42810: gap of unknown length
47071: contig of 4261 bp in length
47151: gap of unknown length
51484: contig of 4333 bp in length
51564: gap of unknown length
55312: contig of 3748 bp in length
55392: gap of unknown length
59360: contig of 3968 bp in length
59440: gap of unknown length
67279: contig of 7839 bp in length
67359: gap of unknown length
71454: contig of 4095 bp in length
71534: gap of unknown length
77682: contig of 6068 bp in length
77682: gap of unknown length
84383: contig of 6701 bp in length
84463: gap of unknown length
92633: contig of 8170 bp in length
92713: gap of unknown length
100750: contig of 8037 bp in length
100830: gap of unknown length
112677: contig of 11847 bp in length
112757: gap of unknown length
121914: contig of 9157 bp in length
121994: gap of unknown length
139643: contig of 17649 bp in length
139723: gap of unknown length
140359: contig of 636 bp in length
140439: gap of unknown length
140439: gap of unknown length
141138: contig of 699 bp in length
141218: gap of unknown length
141882: contig of 764 bp in length
142062: gap of unknown length
143327: contig of 1265 bp in length
143407: gap of unknown length
144194: contig of 787 bp in length
144328: gap of unknown length
144394: gap of unknown length
144775: contig of 746 bp in length
145020: gap of unknown length
145100: gap of unknown length
145588: contig of 488 bp in length
145668: gap of unknown length
146311: contig of 643 bp in length

```

* 146312 146391: gap of unknown length
* 146382 146960: contig of 569 bp in length
* 146961 147040: gap of unknown length
* 147041 147664: contig of 624 bp in length
* 147665 147744: gap of unknown length
* 147745 148437: contig of 693 bp in length
* 148438 148517: gap of unknown length
* 148518 148960: contig of 443 bp in length
* 148961 149040: gap of unknown length
* 149041 149606: contig of 566 bp in length
* 149607 149687: gap of unknown length
* 149688 150570: contig of 884 bp in length
* 150571 150650: gap of unknown length
* 150651 151251: contig of 601 bp in length
* 151252 151331: gap of unknown length
* 151332 152032: contig of 701 bp in length
* 152033 152112: gap of unknown length
* 152113 152660: contig of 548 bp in length
* 152661 152740: gap of unknown length
* 152741 152983: contig of 243 bp in length
* 152984 153063: gap of unknown length
* 153064 153312: contig of 249 bp in length
* 153313 153392: gap of unknown length
* 153393 154241: contig of 849 bp in length
* 154242 154321: gap of unknown length
* 154322 154712: contig of 391 bp in length
* 154713 154792: gap of unknown length
* 154793 155639: contig of 847 bp in length
* 155640 155719: gap of unknown length
* 155720 156286: contig of 567 bp in length
* 156287 156366: gap of unknown length
* 156367 156614: contig of 248 bp in length
* 156615 156694: gap of unknown length
* 156695 157047: contig of 353 bp in length
* 157048 157127: gap of unknown length
* 157128 157560: contig of 433 bp in length
* 157561 157640: gap of unknown length
* 157641 158264: contig of 624 bp in length
* 158265 158344: gap of unknown length
* 158345 158951: contig of 607 bp in length
* 158952 159031: gap of unknown length
* 159032 159561: contig of 530 bp in length
* 159562 159641: gap of unknown length
* 159642 160168: contig of 527 bp in length
* 160169 160248: gap of unknown length
* 160249 160801: contig of 553 bp in length
* 160802 160881: gap of unknown length
* 160882 161518: contig of 637 bp in length.

```

```

FEATURES
Source
1. 161518
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3"
/map="100F1-100F5"
/clone="BACR28N05 (D1183) RPCT-98 28.N.5"
/clone_lib="RPCT-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in

```

```

Query Match 34.6%; Score 18; DB 2; Length 161518;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 21 CATCGCTCTGCTGCTGC 38
Db 86402 CATCGCTCTGCTGCTGC 86419

```

```

RESULT 31
AC013579/c 173093 bp DNA linear HTG 12-MAR-2000
LOCUS AC013579 Homo sapiens clone RP11-3C4, WORKING DRAFT SEQUENCE, 26 unordered
DEFINITION

```

```

ACCESSION
AC013579
VERSION
AC013579.4 GI:7229999
HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 173093)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-3C4
Unpublished
2 (bases 1 to 173093)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barina,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArillano,K., Dewar,K., Domingo,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J.,
Galagan,J.C., Gardyna,S., Grant,G., Hagos,B., Heathord,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karates,A., Klein,J.,
Lehoczky,J., Liew,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLoughlin,J., Meidrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessie,S., Tittrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6884402.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2712
Center clone name: 3 C 4
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158945 bases at least Q40
Consensus quality: 164943 bases at least Q30
Consensus quality: 168064 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 170593; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 26 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 1347: contig of 1347 bp in length
* 1348 1447: gap of 100 bp
* 1448 2769: contig of 1322 bp in length
* 2770 2870 2869: gap of 100 bp
* 2870 4168: contig of 1299 bp in length
* 4169 4268: gap of 100 bp
* 4269 6550: contig of 2282 bp in length
* 6551 6650: gap of 100 bp
* 6651 8819: contig of 2169 bp in length
* 8820 8919: gap of 100 bp
* 8920 10822: contig of 1903 bp in length

```



```
* 10823 10922: gap of 100 bp
* 10923 13176: contig of 2254 bp in length
* 13177 13276: gap of 100 bp
* 13277 15574: contig of 2298 bp in length
* 15575 15674: gap of 100 bp
* 15675 17611: contig of 1937 bp in length
* 17612 17712: gap of 100 bp
* 17712 20922: contig of 3211 bp in length
* 20923 21022: gap of 100 bp
* 21023 24551: contig of 3529 bp in length
* 24552 24652: gap of 100 bp
* 24652 28391: contig of 3739 bp in length
* 28391 28491: gap of 100 bp
* 28491 33813: contig of 5323 bp in length
* 33814 33913: gap of 100 bp
* 33914 37959: contig of 4046 bp in length
* 37960 38059: gap of 100 bp
* 38060 42630: contig of 4571 bp in length
* 42631 42730: gap of 100 bp
* 42731 50918: contig of 8188 bp in length
* 50919 51018: gap of 100 bp
* 51019 59273: contig of 8255 bp in length
* 59274 59373: gap of 100 bp
* 59374 68691: contig of 9318 bp in length
* 68692 68791: gap of 100 bp
* 68792 76416: contig of 7625 bp in length
* 76417 76516: gap of 100 bp
* 76517 83772: contig of 7256 bp in length
* 83773 83872: gap of 100 bp
* 83873 93356: contig of 9484 bp in length
* 93357 93456: gap of 100 bp
* 93457 105377: contig of 11921 bp in length
* 105378 105477: gap of 100 bp
* 105478 118368: contig of 12891 bp in length
* 118369 118468: gap of 100 bp
* 118469 135438: contig of 16970 bp in length
* 135439 135538: gap of 100 bp
* 135539 154849: contig of 19311 bp in length
* 154850 154949: gap of 100 bp
* 154950 173093: contig of 18144 bp in length.
```

FEATURES

```
source
1. 173093
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /clone="RP11-3C4"
   /clone_lib="RP11-11 Human Male BAC"
1. 1347
   /note="assembly_fragment"
1448. 2769
   /note="assembly_fragment"
2870. 4168
   /note="assembly_fragment"
4269. 6550
   /note="assembly_fragment"
6651. 8819
   /note="assembly_fragment"
8920. 10822
   /note="assembly_fragment"
10923. 13176
   /note="assembly_fragment"
13277. 15574
   /note="assembly_fragment"
15675. 17611
   /note="assembly_fragment"
17712. 20922
   /note="assembly_fragment"
21023. 24551
   /note="assembly_fragment"
24652. 28390
   /note="assembly_fragment"
28491. 33813
   /note="assembly_fragment"
```

```
misc_feature 33914..37959
              /note="assembly_fragment"
misc_feature 38060..42630
              /note="assembly_fragment"
misc_feature 42731..50918
              /note="assembly_fragment"
misc_feature 51019..59273
              /note="assembly_fragment"
misc_feature 59374..68691
              /note="assembly_fragment"
misc_feature 68792..76416
              /note="assembly_fragment"
              clone_end:r7
              vector_side:left"
misc_feature 76517..83772
              /note="assembly_fragment"
              clone_end:SP6
              vector_side:right"
misc_feature 83873..93356
              /note="assembly_fragment"
misc_feature 93457..105377
              /note="assembly_fragment"
misc_feature 105478..118368
              /note="assembly_fragment"
misc_feature 118469..135438
              /note="assembly_fragment"
misc_feature 135539..154849
              /note="assembly_fragment"
misc_feature 154950..173093
              /note="assembly_fragment"

BASE COUNT 45233 a 40721 c 40396 g 44224 t 2519 others
ORIGIN
```

```
Query Match 34.6%; Score 18; DB 2; Length 173093;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 3 GCAGTTCCTCAGCCTCT 20
Db 168719 GCAGTTCCTCAGCCTCT 168702
```

```
RESULT 32
AE003845
LOCUS      173874 bp      DNA      linear      INV 14-FEB-2003
DEFINITION Drosophila melanogaster chromosome 4 section 1 of 5 of the complete
            sequence.
ACCESSION  AE003845 AE014135
VERSION    AE003845.4 GI:28380223
KEYWORDS
SOURCE
ORGANISM  Drosophila melanogaster (fruit fly)
            Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 173874)
REFERENCE  1
AUTHORS   Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazer,V.G., Champetier,M., Pfeiffer,B.D.,
            Abail,J.F., Agbayani,A., An,H.J., Andrews-Planckoch,C., Baldwin,D.,
            Ballew,R.M., Baer,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M.,
            Beeson,K.Y., Benos,P.V., Bernan,B.P., Bhandari,D., Bolshakov,S.,
            Borokova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P.,
            Burtis,K.C., Buzam,D.A., Butler,H., Cadeau,E., Center,A.,
            Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B.,
            Davies,P., de Pablos,B., Delcher,A., Deng,Z., Duan,N., Dugan-Rocha,S.,
            Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S.,
            Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,
            Ferreira,S., Fleischmann,W., Fowler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
```

REFERENCE	TITLE	gene	/db_xref="taxon:7227"
REFERENCE	AUTHORS	gene	/chromosome="4"
REFERENCE	AUTHORS	repeat_region	/note="genotype: y[1]: cn[1] bw[1] sp[1]; Rh6[1]"
REFERENCE	AUTHORS	gene	complement(638..1719)
REFERENCE	AUTHORS	repeat_region	/locus_tag="TE20395"
REFERENCE	AUTHORS	gene	complement(638..1719)
REFERENCE	AUTHORS	repeat_region	/map="102A1-102A1"
REFERENCE	AUTHORS	gene	/transposon="pagg1ne[1]471"
REFERENCE	AUTHORS	repeat_region	/db_xref="FLYBASE:FBF10020395"
REFERENCE	AUTHORS	gene	complement(12554..4264)
REFERENCE	AUTHORS	repeat_region	/locus_tag="TE20396"
REFERENCE	AUTHORS	gene	complement(12554..4264)
REFERENCE	AUTHORS	repeat_region	/map="102A1-102A1"
REFERENCE	AUTHORS	gene	/transposon="Rtic[1]1472"
REFERENCE	AUTHORS	repeat_region	/db_xref="FLYBASE:FBF10020396"
REFERENCE	AUTHORS	gene	4885..11664
REFERENCE	AUTHORS	repeat_region	/locus_tag="TE20397"
REFERENCE	AUTHORS	gene	4885..11664
REFERENCE	AUTHORS	repeat_region	/map="102A1-102A1"
REFERENCE	AUTHORS	gene	/transposon="GATE[1]1473"
REFERENCE	AUTHORS	repeat_region	/db_xref="FLYBASE:FBF10020397"
REFERENCE	AUTHORS	gene	complement(11691..12255)
REFERENCE	AUTHORS	repeat_region	/locus_tag="TE20398"
REFERENCE	AUTHORS	gene	complement(11691..12255)
REFERENCE	AUTHORS	repeat_region	/map="102A1-102A1"
REFERENCE	AUTHORS	gene	/transposon="GATE[1]1474"
REFERENCE	AUTHORS	repeat_region	/db_xref="FLYBASE:FBF10020398"
REFERENCE	AUTHORS	gene	12291..113244
REFERENCE	AUTHORS	repeat_region	/locus_tag="TE20399"
REFERENCE	AUTHORS	gene	12291..113244
REFERENCE	AUTHORS	repeat_region	/map="102A1-102A1"
REFERENCE	AUTHORS	gene	/transposon="GATE[1]1475"
REFERENCE	AUTHORS	repeat_region	/db_xref="FLYBASE:FBF10020399"
REFERENCE	AUTHORS	gene	complement(113288..113761)
REFERENCE	AUTHORS	repeat_region	/locus_tag="TE20400"
REFERENCE	AUTHORS	gene	complement(113288..113761)
REFERENCE	AUTHORS	repeat_region	/map="102A1-102A1"
REFERENCE	AUTHORS	gene	/transposon="1360[1]1476"
REFERENCE	AUTHORS	repeat_region	/db_xref="FLYBASE:FBF10020400"
REFERENCE	AUTHORS	gene	complement(117702..18272)
REFERENCE	AUTHORS	repeat_region	/locus_tag="TE20401"
REFERENCE	AUTHORS	gene	complement(117702..18272)
REFERENCE	AUTHORS	repeat_region	/map="102A1-102A1"
REFERENCE	AUTHORS	gene	/transposon="Rtlb[1]1477"
REFERENCE	AUTHORS	repeat_region	/db_xref="FLYBASE:FBF10020401"
REFERENCE	AUTHORS	gene	complement(122335..23205)
REFERENCE	AUTHORS	repeat_region	/locus_tag="CG32013"
REFERENCE	AUTHORS	gene	/note="Last curated on Thu Apr 25 15:01:21 PDT 2002"
REFERENCE	AUTHORS	repeat_region	/map="102A1-102A1"
REFERENCE	AUTHORS	gene	/db_xref="FLYBASE:FBgn0052013"
REFERENCE	AUTHORS	repeat_region	/complement(join(22335..22528,22617..23205))
REFERENCE	AUTHORS	gene	/locus_tag="CG32013"
REFERENCE	AUTHORS	repeat_region	/note="CG32013 gene product"
REFERENCE	AUTHORS	gene	/codon_start=1
REFERENCE	AUTHORS	repeat_region	/product="CG32013-PA"
REFERENCE	AUTHORS	gene	/protein_id="AAN06536.1"
REFERENCE	AUTHORS	repeat_region	/db_xref="GI:22759415"
REFERENCE	AUTHORS	gene	/db_xref="FLYBASE:FBgn0052013"
REFERENCE	AUTHORS	repeat_region	/translation="MGAAADAAPGTTTAAALVVAAPHLVRASTPSSPGLLPGPS
REFERENCE	AUTHORS	gene	CHRRPPISTVTTWPTCATCGISHAASCLSRGSHSGVSGASSASIDAGHVHTDC
REFERENCE	AUTHORS	repeat_region	SGACAGVIRGGVGGDCVDCGRCGGGCGGCGGAETTPAIBALAIRAVSSSD
REFERENCE	AUTHORS	gene	TGCATVPPSSSSLSERVIAAAPVFAIGMSVFPCPR"
REFERENCE	AUTHORS	repeat_region	24068..22621
REFERENCE	AUTHORS	gene	/locus_tag="CG17923"
REFERENCE	AUTHORS	repeat_region	/note="Last curated on Thu Apr 25 15:33:46 PDT 2002"
REFERENCE	AUTHORS	gene	/map="102A1-102A1"
REFERENCE	AUTHORS	repeat_region	/db_xref="FLYBASE:FBgn0040037"
REFERENCE	AUTHORS	gene	join(24068..24477,24979..25153,25218..25450,25501..25621)

CDS

```

/locus_tag="CG17923"
/product="CG17923-RA"
/db_xref="FLYBASE:FBgn0040037"
join(124134..24477,24979..25153,25218..25450,25501..25621)
/locus_tag="CG17923"
/note="CG17923 gene product"
/codon_start=1
/product="CG17923-PA"
/protein_id="AAN06537.1"
/db_xref="GI:22759416"
/db_xref="FLYBASE:FBgn0040037"
/translation="MGISGSDVSKQADHILDDNFASIVGIEGRILFDNLKKSIA
YTLTNDLPIVPLFFVIFDIPDLAATIALICIDIGTDLPLISLAVEKEDIDAMR
PRDPEDRLVNNKLLIMAYLIQIVTACFTFPAIMAKHGFPSRLKGIREDMSK
NVEDLEDGGEQMTYRERKVLKLEGTAGTGFVSIVTQVDFDLICRRNSIIQCGMGN
HYLNFALVEFIACLLCVPFVEKTLNMYIKFIWIAVAFGLLIFFEDSRFLI
RRNPGWVEQERY"
complement(26482..34110)
/locus_tag="CG32011"
/note="last curated on Thu Aug 29 10:39:29 PDT 2002"
/map="102A1-102A1"
/db_xref="FLYBASE:FBgn0052011"
complement(join(26482..26667,27167..27349,28371..28609,
28966..29301,29356..30391,30551..31625,31703..33391,
33949..34110))

```

Query Match 34.6% Score 18; DB 3; Length 173874;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 CATCGCTCTGCTGCTGC 38
12415 CATCGCTCTGCTGCTGC 12432

RESULT 33
AC007587 185916 bp DNA linear INV 14-DEC-2001
LOCUS BACR02D22, complete sequence.
AC007587
DEFINITION
AC007587.11 GI:17737031
HTG.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 185916)
Celinker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amandides,P.G., Brannon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dreisner,D., Fartan,D.,
Fertile,S., Frise,E., Galle,R.F., Garey,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hoslin,D., Howland,T.J.,
Idegman,C., Jalili,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Neilson,C., Nelson,K.A., Nunco,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phuanavong,S., Piltman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svitskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 2R, region 41C-41D
Unpublished (1998)
2 (bases 1 to 185916)
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Buenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Fartan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Shit,E.,
Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and

TITLE
JOURNAL
COMMENT

Rubin,G.M.
Direct Submission
Submitted (19-MAY-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Dec 14, 2001 this sequence version replaced gi:7264774.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.
Location/Qualifiers
1..185916
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="41C-41D"
/clone_1b="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcorI in
pBAC3.6)"
pBAC3.6")

BASE COUNT 56016 a 37375 c 36026 g 56499 t
ORIGIN

Query Match 34.6% Score 18; DB 3; Length 185916;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 CATCGCTCTGCTGCTGC 38
29892 CATCGCTCTGCTGCTGC 29909

RESULT 34
AC008333/C 187223 bp DNA linear HTG 15-MAR-2002
LOCUS AC008333, complete sequence.
AC008333
DEFINITION
AC008333.22 GI:19482330
HTG; HTGS PHASE1.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 187223)
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Buenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Fartan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svitskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 187223)
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Buenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Fartan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Shit,E.,
Svitskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and

TITLE
JOURNAL
REFERENCE
AUTHORS
Rubin,G.M.
Direct Submission
Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
3 (bases 1 to 187223)
Celniker,S.E., Aghayani,A., Arcaine,T.T., Baxter,E., Blazej,R.G.,
Brenneise,C., Champetier,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farnham,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karta,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Paciorek,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Swirekas,R.R., Tan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Direct Submission
Submitted (15-MAR-2002) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 15, 2002 this sequence version replaced gi:7264781.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to dbgap@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 117 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
400 399: contig of 399 bp in length
479: gap of unknown length
480 1585: contig of 1106 bp in length
1586 1665: gap of unknown length
1666 3031: contig of 1366 bp in length
3032 3111: gap of unknown length
3112 4483: contig of 1372 bp in length
4484 4563: gap of unknown length
4564 5967: contig of 1404 bp in length
5968 6047: gap of unknown length
6048 8158: contig of 2111 bp in length
8159 8238: gap of unknown length
8239 9368: contig of 1130 bp in length
9369 9448: gap of unknown length
9449 11398: contig of 1950 bp in length
11399 11478: gap of unknown length
11479 12823: contig of 1345 bp in length
12824 12903: gap of unknown length
12904 15331: contig of 2428 bp in length
15332 15411: gap of unknown length
15412 16843: contig of 1432 bp in length
16844 16923: gap of unknown length
16924 19239: contig of 2316 bp in length
19240 19319: gap of unknown length
19320 20191: contig of 872 bp in length
20192 20271: gap of unknown length
20272 23177: contig of 2906 bp in length
23178 23257: gap of unknown length
23258 24861: contig of 1604 bp in length
24862 24941: gap of unknown length
24942 28264: contig of 3323 bp in length
28265 28344: gap of unknown length
28345 32799: contig of 4455 bp in length
32800 32879: gap of unknown length
32880 44930: contig of 12051 bp in length
44931 45010: gap of unknown length
45011 59268: contig of 14258 bp in length
59269 59348: gap of unknown length
59349 73716: contig of 14368 bp in length
73717 73796: gap of unknown length
73797 85947: contig of 12151 bp in length
85948 86027: gap of unknown length
86028 107333: contig of 21306 bp in length

107334 107413: gap of unknown length
107414 116139: contig of 8726 bp in length
116140 116219: gap of unknown length
116220 130488: contig of 14269 bp in length
130489 130568: gap of unknown length
130569 130881: contig of 313 bp in length
130882 130961: gap of unknown length
130962 131434: contig of 473 bp in length
131435 131514: gap of unknown length
131515 132452: contig of 938 bp in length
132453 132532: gap of unknown length
132533 133533: contig of 1001 bp in length
133534 133613: gap of unknown length
133614 134016: contig of 403 bp in length
134017 134096: gap of unknown length
134097 134556: contig of 460 bp in length
134557 134636: gap of unknown length
134637 135518: contig of 882 bp in length
135519 135598: gap of unknown length
135599 136290: contig of 692 bp in length
136291 136370: gap of unknown length
136371 136924: contig of 554 bp in length
136925 137004: gap of unknown length
137005 137715: contig of 711 bp in length
137716 137795: gap of unknown length
137796 138094: contig of 299 bp in length
138095 138174: gap of unknown length
138175 139004: contig of 830 bp in length
139005 139084: gap of unknown length
139085 139649: contig of 565 bp in length
139650 139729: gap of unknown length
139730 140353: contig of 624 bp in length
140354 140433: gap of unknown length
140434 140639: contig of 206 bp in length
140640 140719: gap of unknown length
140720 141001: contig of 282 bp in length
141002 141081: gap of unknown length
141082 142004: contig of 923 bp in length
142005 142084: gap of unknown length
142085 142643: contig of 559 bp in length
142644 142723: gap of unknown length
142724 143061: contig of 338 bp in length
143062 143141: gap of unknown length
143142 143398: contig of 257 bp in length
143399 143478: gap of unknown length
143479 144700: contig of 1222 bp in length
144701 144780: gap of unknown length
144781 145010: contig of 230 bp in length
145011 145090: gap of unknown length
145091 145783: contig of 693 bp in length
145784 145863: gap of unknown length
145864 146308: contig of 445 bp in length
146309 146388: gap of unknown length
146389 147049: contig of 661 bp in length
147050 147129: gap of unknown length
147130 147551: contig of 422 bp in length
147552 147631: gap of unknown length
147632 148164: contig of 533 bp in length
148165 148244: gap of unknown length
148245 148503: contig of 259 bp in length
148504 148583: gap of unknown length
148584 149079: contig of 496 bp in length
149080 149159: gap of unknown length
149160 150009: contig of 850 bp in length
150010 150089: gap of unknown length
150090 151007: contig of 918 bp in length
151008 151087: gap of unknown length
151088 151607: contig of 520 bp in length
151608 151687: gap of unknown length
151688 152125: contig of 438 bp in length
152126 152205: gap of unknown length
152206 152550: contig of 345 bp in length
152551 152630: gap of unknown length

```

* 152631 153442: contig of 712 bp in length
* 153343 153422: gap of unknown length
* 153423 154116: contig of 694 bp in length
* 154117 154196: gap of unknown length
* 154197 154442: contig of 446 bp in length
* 154643 154722: gap of unknown length
* 154723 155046: contig of 324 bp in length
* 155047 155126: gap of unknown length
* 155127 156443: contig of 1317 bp in length
* 156444 156523: gap of unknown length
* 156524 157098: contig of 575 bp in length
* 157099 157178: gap of unknown length
* 157179 157983: contig of 805 bp in length
* 157984 158063: gap of unknown length
* 158064 159681: contig of 1618 bp in length
* 159682 159761: gap of unknown length
* 159762 160166: contig of 405 bp in length
* 160167 160246: gap of unknown length
* 160247 160579: contig of 331 bp in length
* 160580 160659: gap of unknown length
* 160660 160963: contig of 304 bp in length
* 160964 161043: gap of unknown length

```

```

Query Match      34.6%; Score 18; DB 2; Length 187223;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 21 CATGCTCTCGTCGCTGC 38
Db 129760 CATGCTCTCGTCGCTGC 129743

```

```

RESULT 35
AC016393/c
LOCUS
DEFINITION Homo sapiens chromosome 12 clone RP11-199C7 map 12, *** SEQUENCING
IN PROGRESS ***, 91 unordered pieces.
ACCESSION AC016393
VERSION AC016393.3
KEYWORDS HTG: HTGS PHASE1
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 188032)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 12, clone RP11-199C7
Unpublished
2 (bases 1 to 188032)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,K., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karats,A., Klein,J.,
Lehoczky,J., Lieu,C., Johnson,R., Jones,C., Kann,L., Karats,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McGraw,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,S., Subramanian,A., Talamas,J.,
Teefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 188032)

```

TITLE JOURNAL COMMENT

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karats,A.,
Klein,J., Lakoque,K., Lamazares,R., Lander,T., Lehoczky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McGraw,P., McGurk,A., McKernan,K., McPheters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,S., Subramanian,A., Talamas,J.,
Teefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6759157.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4953
Center clone name: 199_C_7

* NOTE: This is a 'working draft' sequence. It currently
* consists of 91 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1263: contig of 1263 bp in length
1264 1363: gap of 100 bp
1364 2365: contig of 1002 bp in length
2366 2465: gap of 100 bp
2466 3982: contig of 1517 bp in length
3983 4082: gap of 100 bp
4083 5089: contig of 1007 bp in length
5090 5189: gap of 100 bp
5190 6231: contig of 1042 bp in length
6232 6331: gap of 100 bp
6332 7580: contig of 1249 bp in length
7581 7680: gap of 100 bp
7681 8972: contig of 1292 bp in length
8973 9072: gap of 100 bp
9073 10251: contig of 1179 bp in length
10252 10351: gap of 100 bp
10352 11684: contig of 1333 bp in length
11685 11784: gap of 100 bp
11785 12954: contig of 1170 bp in length
12955 13054: gap of 100 bp
13055 14166: contig of 1112 bp in length
14167 14265: gap of 100 bp
14267 15381: contig of 1115 bp in length
15382 15481: gap of 100 bp
15482 16485: contig of 1004 bp in length
16486 16585: gap of 100 bp
16586 17950: contig of 1365 bp in length
17951 18050: gap of 100 bp
18051 18062: contig of 12 bp in length
18063 18162: gap of 100 bp
18163 18591: contig of 1429 bp in length
18592 20869: gap of 100 bp
20870 20969: contig of 1178 bp in length
20969: gap of 100 bp

```

```

20970 22151: contig of 1182 bp in length
* 22152 22251: gap of 100 bp
* 22252 23495: contig of 1244 bp in length
* 23496 23595: gap of 100 bp
* 23596 24843: contig of 1248 bp in length
* 24844 24943: gap of 100 bp
* 24944 26155: contig of 1212 bp in length
* 26156 26255: gap of 100 bp
* 26256 27991: contig of 1736 bp in length
* 27992 28091: gap of 100 bp
* 28092 29183: contig of 1092 bp in length
* 29184 29283: gap of 100 bp
* 29284 30293: contig of 1010 bp in length
* 30294 30393: gap of 100 bp
* 30394 31466: contig of 1073 bp in length
* 31467 31566: gap of 100 bp
* 31567 32928: contig of 1362 bp in length
* 32929 33028: gap of 100 bp
* 33029 34776: contig of 1748 bp in length
* 34777 34876: gap of 100 bp
* 34877 36265: contig of 1389 bp in length
* 36266 36365: gap of 100 bp
* 36366 37605: contig of 1240 bp in length
* 37606 37705: gap of 100 bp
* 37706 38974: contig of 1269 bp in length
* 38975 39074: gap of 100 bp
* 39075 40313: contig of 1239 bp in length
* 40314 40413: gap of 100 bp
* 40414 41544: contig of 1131 bp in length
* 41545 41644: gap of 100 bp
* 41645 43458: contig of 1814 bp in length
* 43459 43558: gap of 100 bp
* 43559 45089: contig of 1531 bp in length
* 45090 45189: gap of 100 bp
* 45190 46414: contig of 1225 bp in length
* 46415 46514: gap of 100 bp
* 46515 47776: contig of 1262 bp in length
* 47777 47876: gap of 100 bp
* 47877 49306: contig of 1430 bp in length
* 49307 49406: gap of 100 bp
* 49407 50583: contig of 1177 bp in length
* 50584 50683: gap of 100 bp
* 50684 52614: contig of 1931 bp in length
* 52615 52714: gap of 100 bp
* 52715 54063: contig of 1349 bp in length
* 54064 54163: gap of 100 bp
* 54164 55229: contig of 1066 bp in length
* 55230 55329: gap of 100 bp
* 55330 56490: contig of 1161 bp in length
* 56491 56590: gap of 100 bp
* 56591 57831: contig of 1241 bp in length
* 57832 57931: gap of 100 bp
* 57932 59583: contig of 1652 bp in length
* 59584 59683: gap of 100 bp
* 59684 61686: contig of 2003 bp in length
* 61687 61786: gap of 100 bp
* 61787 63126: contig of 1340 bp in length
* 63127 63226: gap of 100 bp
* 63227 64917: contig of 1691 bp in length
* 64918 65017: gap of 100 bp
* 65018 66634: contig of 1617 bp in length
* 66635 66734: gap of 100 bp
* 66735 68500: contig of 1766 bp in length
* 68501 68600: gap of 100 bp
* 68601 70266: contig of 1666 bp in length
* 70267 70366: gap of 100 bp
* 70367 72119: contig of 1753 bp in length
* 72120 72219: gap of 100 bp
* 72220 74200: contig of 1981 bp in length
* 74201 74300: gap of 100 bp
* 74301 76382: contig of 2082 bp in length
* 76383 76482: gap of 100 bp
* 76483 78034: contig of 1552 bp in length

```

```

Query Match      34.6%  Score 18; DB 2; Length 188032;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 AGTTCTCTCAGCCTCTCA 22
Db 116725 AGTTCTCTCAGCCTCTCA 116708

```

```

RESULT 36
AC013254
LOCUS
DEFINITION
Drosophila melanogaster chromosome 2 clone BACR25010 (D1173)
PC1-98.25.O.10 map 40D-40D strain Y; cn bw sp. *** SEQUENCING IN
PROGRESS ***; 184 unordered pieces.

```

```

ACCESSION
AC013254
VERSION
AC013254.6 GI:7259653
KEYWORDS
HTG; HTGS PHASE1.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 189893)

```

```

REFERENCE
AUTHORS
Celiker,S.E., Aghayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champagne,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karta,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Swirek,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 189893)

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Butenhoff,C., Champagne,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karta,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Swirek,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Direct Submission
Submitted (05-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, WS 64-121, Berkeley, CA 94720, USA
On Mar 17, 2000 this sequence version replaced gi:6984327.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to hdgeff@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.

```

```

COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 184 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 588: contig of 588 bp in length
* 589 668: gap of unknown length
* 1063: contig of 395 bp in length
* 1064 1143: gap of unknown length
* 1144 1659: contig of 516 bp in length
* 1660 1739: gap of unknown length
* 1740 2258: contig of 519 bp in length
* 2259 2338: gap of unknown length
* 2339 2533: contig of 295 bp in length
* 2634 2713: gap of unknown length
* 2714 3238: contig of 525 bp in length
* 3239 3318: gap of unknown length
* 3319 3832: contig of 514 bp in length
* 3833 3912: gap of unknown length
* 3913 4239: contig of 327 bp in length
* 4240 4320 4912: contig of 593 bp in length
* 4913 4992: gap of unknown length
* 4993 5610: contig of 618 bp in length
* 5611 5690: gap of unknown length
* 5691 5976: contig of 286 bp in length
* 5977 6056: gap of unknown length
* 6057 6580: contig of 524 bp in length
* 6581 7196: contig of 516 bp in length
* 7197 7276: gap of unknown length
* 7277 7861: contig of 585 bp in length
* 7862 7941: gap of unknown length
* 7942 9001: contig of 1060 bp in length
* 9002 9684: contig of 603 bp in length
* 9685 9764: gap of unknown length
* 9765 10668: contig of 904 bp in length
* 10669 10748: gap of unknown length
* 10749 11441: contig of 693 bp in length
* 11442 11521: gap of unknown length
* 11522 12133: contig of 612 bp in length
* 12134 12213: gap of unknown length
* 12214 13064: contig of 851 bp in length
* 13065 13144: gap of unknown length
* 13145 13935: contig of 791 bp in length
* 13936 14015: gap of unknown length
* 14016 14921: contig of 906 bp in length
* 14922 15001: gap of unknown length
* 15002 16065: contig of 1064 bp in length
* 16066 16145: gap of unknown length
* 16146 16953: contig of 808 bp in length
* 16954 17033: gap of unknown length
* 17034 17569: contig of 536 bp in length
* 17570 17649: gap of unknown length
* 17650 18097: contig of 448 bp in length
* 18098 18177: gap of unknown length
* 18178 18940: contig of 763 bp in length
* 18941 19020: gap of unknown length
* 19021 19247: contig of 227 bp in length
* 19248 19327: gap of unknown length
* 19328 20043: contig of 716 bp in length
* 20044 20123: gap of unknown length
* 20124 20734: contig of 611 bp in length
* 20735 20814: gap of unknown length
* 20815 21151: contig of 337 bp in length
* 21152 21231: gap of unknown length
* 21232 21885: contig of 654 bp in length
* 21886 21965: gap of unknown length
* 21966 23030: contig of 1065 bp in length
* 23031 23110: gap of unknown length
* 23111 24601: contig of 1491 bp in length
* 24602 24681: gap of unknown length
* 24682 25250: contig of 569 bp in length
* 25251 25330: gap of unknown length

* 25331 26463: contig of 1133 bp in length
* 26464 26543: gap of unknown length
* 26544 26937: contig of 394 bp in length
* 26938 27017: gap of unknown length
* 27018 28151: contig of 1134 bp in length
* 28152 28231: gap of unknown length
* 28232 29547: contig of 1316 bp in length
* 29548 29627: gap of unknown length
* 29628 30471: contig of 844 bp in length
* 30472 30551: gap of unknown length
* 30552 32369: contig of 1818 bp in length
* 32370 32449: gap of unknown length
* 32450 33789: contig of 1340 bp in length
* 33790 33869: gap of unknown length
* 33870 35224: contig of 1355 bp in length
* 35225 35204: gap of unknown length
* 35226 35305: contig of 1249 bp in length
* 35306 36553: gap of unknown length
* 36554 36633: gap of unknown length
* 36634 38274: contig of 1641 bp in length
* 38275 38354: gap of unknown length
* 38355 38889: gap of unknown length
* 38890 39969: gap of unknown length
* 39970 41732: contig of 1762 bp in length
* 41732 41812 43829: contig of 2018 bp in length
* 43830 43910 46534: contig of 2725 bp in length
* 46535 46714: gap of unknown length
* 46715 48565: contig of 1851 bp in length
* 48566 48645: gap of unknown length
* 48646 50553: contig of 1908 bp in length
* 50554 50633: gap of unknown length
* 50634 53539: contig of 3006 bp in length
* 53540 53719: gap of unknown length
* 53720 55306: contig of 1587 bp in length
* 55307 61387: gap of unknown length
* 61388 61467: gap of unknown length
* 61468 66422: contig of 4955 bp in length
* 66423 66502: gap of unknown length
* 66503 72908: contig of 6406 bp in length
* 72909 72988: gap of unknown length
* 72989 79928: contig of 6940 bp in length
* 79929 80009: gap of unknown length
* 80010 88702: contig of 8654 bp in length
* 88703 88783 99255: contig of 10473 bp in length
* 88784 99335: gap of unknown length
* 99336 99315 99914: contig of 579 bp in length
* 99915 99994: gap of unknown length
* 99995 100377: contig of 383 bp in length
* 100378 100457: gap of unknown length
* 100458 100944: contig of 487 bp in length
* 100945 101024: gap of unknown length
* 101025 101605: contig of 581 bp in length
* 101606 101685: gap of unknown length
* 101686 102540: contig of 855 bp in length
* 102541 102620: gap of unknown length
* 102621 103304: contig of 684 bp in length
* 103305 103384: gap of unknown length
* 103385 103818: contig of 434 bp in length
* 103819 103898: gap of unknown length
* 103899 104685: contig of 787 bp in length
* 104686 104765: gap of unknown length
* 104766 105637: contig of 872 bp in length
* 105638 105718: gap of unknown length
* 105719 106017: contig of 300 bp in length
* 106018 106097: gap of unknown length
* 106098 106348: contig of 251 bp in length
* 106349 106428: gap of unknown length
* 106429 107046: contig of 618 bp in length
* 107047 107126: gap of unknown length
* 107127 107624: contig of 498 bp in length

```
* 107625 107704: gap of unknown length
* 107705 108652: contig of 958 bp in length
* 108663 108742: gap of unknown length
* 108743 109478: contig of 736 bp in length
* 109479 109558: gap of unknown length
* 109559 110172: contig of 614 bp in length
* 110173 110252: gap of unknown length
* 110253 110986: contig of 734 bp in length

Query Match      34.6%; Score 18; DB 2; Length 189893;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 CATCGCTCTGCTGCTGC 38
DB 64905 CATCGCTCTGCTGCTGC 64922

RESULT 37
AC013254/c
LOCUS
DEFINITION
AC013254 189893 bp DNA linear HTG 17-MAR-2000
Drosophila melanogaster chromosome 2 clone BACR25010 (D1173)
PFC1-98 25.010 map 40D-40D strain Y; cn bw sp, *** SEQUENCING IN
PROCESS ***, 184 unordered pieces.
AC013254
AC013254.6 GI:7259653
HTG; HTGS_PHASE1.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 189893)
Celtnker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoft,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacled,J.M., Park,S., Pfeiffer,B.,
Richard,S., Sethi,H., Swirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 189893)
Celtnker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoft,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacled,J.M., Park,S., Pfeiffer,B.,
Richard,S., Sethi,H., Swirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Direct Submission
Submitted (05-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 17, 2000 this sequence version replaced gi:6984327.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 184 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 588: contig of 588 bp in length
* 589 568: gap of unknown length
* 669 1063: contig of 395 bp in length
* 1064 1143: gap of unknown length
```



```

* 29548 29627: gap of unknown length
* 29628 30471: contig of 844 bp in length
* 30472 30551: gap of unknown length
* 30552 32369: contig of 1818 bp in length
* 32370 32449: gap of unknown length
* 32450 33789: contig of 1340 bp in length
* 33790 33869: gap of unknown length
* 33870 35224: contig of 1355 bp in length
* 35225 35304: gap of unknown length
* 35305 36553: contig of 1249 bp in length
* 36554 36633: gap of unknown length
* 36634 38274: contig of 1641 bp in length
* 38275 38354: gap of unknown length
* 38355 39889: contig of 1535 bp in length
* 39890 39969: gap of unknown length
* 39970 41731: contig of 1762 bp in length
* 41732 41811: gap of unknown length
* 41812 43829: contig of 2018 bp in length
* 43830 43909: gap of unknown length
* 46634 46714: contig of 2725 bp in length
* 46715 48565: contig of 1851 bp in length
* 48566 48645: gap of unknown length
* 48646 50553: contig of 1908 bp in length
* 50554 50633: gap of unknown length
* 50634 53639: contig of 3006 bp in length
* 53640 53719: gap of unknown length
* 53720 55306: contig of 1587 bp in length
* 55307 55386: gap of unknown length
* 55387 61387: contig of 6001 bp in length
* 61388 61467: gap of unknown length
* 61468 66422: contig of 4955 bp in length
* 66423 66502: gap of unknown length
* 66503 72908: contig of 6406 bp in length
* 72909 72988: gap of unknown length
* 72989 79928: contig of 6940 bp in length
* 79929 80008: gap of unknown length
* 80009 88702: contig of 8694 bp in length
* 88703 88782: gap of unknown length
* 88783 99255: contig of 10473 bp in length
* 99256 99335: gap of unknown length
* 99336 99914: contig of 579 bp in length
* 99915 99994: gap of unknown length
* 99995 100377: contig of 383 bp in length
* 100378 100457: gap of unknown length
* 100458 100944: contig of 487 bp in length
* 100945 101024: gap of unknown length
* 101025 101605: contig of 581 bp in length
* 101606 101885: gap of unknown length
* 101886 102540: contig of 855 bp in length
* 102541 102620: gap of unknown length
* 102621 103304: contig of 684 bp in length
* 103305 103384: gap of unknown length
* 103385 103818: contig of 433 bp in length
* 103819 103898: gap of unknown length
* 103899 104685: contig of 787 bp in length
* 104686 104765: gap of unknown length
* 104766 105637: contig of 872 bp in length
* 105638 105717: gap of unknown length
* 105718 106017: contig of 300 bp in length
* 106018 106097: gap of unknown length
* 106098 106348: contig of 251 bp in length
* 106349 106428: gap of unknown length
* 106429 107046: contig of 618 bp in length
* 107047 107126: gap of unknown length
* 107127 107624: contig of 498 bp in length
* 107625 107704: gap of unknown length
* 107705 108662: contig of 958 bp in length
* 108663 108742: gap of unknown length
* 108743 109478: contig of 736 bp in length
* 109479 109558: gap of unknown length
* 109559 110172: contig of 614 bp in length
* 110173 110252: gap of unknown length

```

```

* 110253 110986: contig of 734 bp in length

Query Match      34.6% Score 18; DB 2; Length 189893;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      21 CATCGCTCTGCTGCTGC 38
|||||
Db      76500 CATCGCTCTGCTGCTGC 76483

```

```

RESULT 38
AC007415
LOCUS
DEFINITION
AC007415      198282 bp      DNA      linear      HTG 17-MAR-2000
Drosophila melanogaster chromosome 2 clone BACR05A24 (ID603) RPC1-98
05.A.24 map 41C-41D strain Y; cn bw sp. *** SEQUENCING IN PROGRESS
***.103 unordered pieces.

```

```

ACCESSION
AC007415
VERSION
AC007415.11 GI:7259656
KEYWORDS
HTG; HTGS PHASE1
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

```

```

REFERENCE
1 (bases 1 to 198282)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Lomcan,M.A., Mazda,P.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished

```

```

TITLE
2 (bases 1 to 198282)
JOURNAL
REFERENCE
AUTHORS
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomcan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Shir,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zietan,L.L. and
Rubin,G.M.

```

```

TITLE
Direct Submission
JOURNAL
Submitted (28-Apr-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 17, 2000 this sequence version replaced gi:7159347.
COMMENT
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgs@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 103 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 926: contig of 926 bp in length
927 1006: gap of unknown length
1007 1301: contig of 295 bp in length
1302 1381: gap of unknown length
1382 1860: contig of 479 bp in length
1861 1940: gap of unknown length
1941 2403: contig of 463 bp in length
2404 2483: gap of unknown length
2484 2895: contig of 412 bp in length
2896 2975: gap of unknown length

```

2976 3478: contig of 503 bp in length
3479 3558: gap of unknown length
3559 4034: contig of 476 bp in length
4035 4114: gap of unknown length
4115 4788: contig of 674 bp in length
4789 4868: gap of unknown length
4869 5549: contig of 681 bp in length
5550 5629: gap of unknown length
5630 6443: contig of 814 bp in length
6444 6523: gap of unknown length
6524 7697: contig of 1174 bp in length
7698 7777: gap of unknown length
7778 9243: contig of 1466 bp in length
9244 9323: gap of unknown length
9324 12162: contig of 2839 bp in length
12163 12242: gap of unknown length
12243 14263: contig of 2021 bp in length
14264 14343: gap of unknown length
14344 16104: contig of 1761 bp in length
16105 16184: gap of unknown length
16185 18228: contig of 2044 bp in length
18229 18308: gap of unknown length
20874: contig of 2566 bp in length
20954: gap of unknown length
20955 24951: contig of 3997 bp in length
24952 25031: gap of unknown length
25032 28493: contig of 3462 bp in length
28494 28573: gap of unknown length
28574 31026: contig of 2453 bp in length
31027 31106: gap of unknown length
31107 35546: contig of 4440 bp in length
35547 35626: gap of unknown length
35627 40849: contig of 5223 bp in length
40850 40929: gap of unknown length
40930 48761: contig of 7832 bp in length
48762 48841: gap of unknown length
48842 52556: contig of 3815 bp in length
52557 52736: gap of unknown length
52737 58188: contig of 5452 bp in length
58189 58268: gap of unknown length
62896: contig of 4628 bp in length
62897 62976: gap of unknown length
62977 72175: contig of 9199 bp in length
72176 72255: gap of unknown length
72256 86109: contig of 13854 bp in length
86110 86189: gap of unknown length
94772: contig of 8583 bp in length
94773 94852: gap of unknown length
94853 112088: contig of 17236 bp in length
112089 112168: gap of unknown length
112169 128075: contig of 15907 bp in length
128076 128155: gap of unknown length
128156 146641: contig of 18486 bp in length
146642 146721: gap of unknown length
147805: contig of 1084 bp in length
147806 147885: gap of unknown length
147886 148190: contig of 305 bp in length
148191 148270: gap of unknown length
148271 149254: contig of 984 bp in length
149255 149334: gap of unknown length
149335 150046: contig of 712 bp in length
150047 150126: gap of unknown length
150127 151028: contig of 902 bp in length
151029 151108: gap of unknown length
151109 151716: contig of 608 bp in length
151717 151796: gap of unknown length
151797 152455: contig of 659 bp in length
152456 152535: gap of unknown length
152536 153204: contig of 669 bp in length
153205 153284: gap of unknown length
153285 153938: contig of 654 bp in length
153939 154018: gap of unknown length
154019 154280: contig of 262 bp in length

154281 154360: gap of unknown length
154361 155656: contig of 1296 bp in length
155657 155736: gap of unknown length
155737 156329: contig of 592 bp in length
156329 156409: gap of unknown length
156409 157266: contig of 858 bp in length
157267 157346: gap of unknown length
157347 157827: contig of 481 bp in length
157828 157907: gap of unknown length
157908 157988: contig of 1066 bp in length
157989 158994: gap of unknown length
158994 159073: gap of unknown length
159074 159316: contig of 243 bp in length
159317 159396: gap of unknown length
159397 159756: contig of 360 bp in length
159757 159836: gap of unknown length
159837 160697: contig of 861 bp in length
160698 160777: gap of unknown length
160778 161496: contig of 719 bp in length
161497 161576: gap of unknown length
161577 162155: contig of 579 bp in length
162156 162235: gap of unknown length
162236 162542: contig of 307 bp in length
162543 162582: gap of unknown length
162583 162964: contig of 342 bp in length
162965 163044: gap of unknown length
163045 163544: contig of 500 bp in length
163545 163624: gap of unknown length
163625 165092: contig of 1468 bp in length
165093 165172: gap of unknown length
165173 165866: contig of 694 bp in length
165867 165946: gap of unknown length
165947 166183: contig of 237 bp in length
166184 166263: gap of unknown length
166264 167263: contig of 1000 bp in length
167264 167343: gap of unknown length
167344 167883: contig of 540 bp in length
167884 167963: gap of unknown length
167964 168338: contig of 375 bp in length
168339 168418: gap of unknown length
168419 168719: contig of 301 bp in length
168720 168799: gap of unknown length
168800 169480: contig of 681 bp in length
169481 169560: gap of unknown length
169561 170667: contig of 507 bp in length
170668 170147: gap of unknown length
170148 170904: contig of 757 bp in length
170905 170984: gap of unknown length
170985 171769: contig of 785 bp in length
171770 171849: gap of unknown length
171849 172398: contig of 549 bp in length
172399 172478: gap of unknown length
172479 173200: contig of 722 bp in length
173201 173280: gap of unknown length
173281 173833: contig of 553 bp in length
173834 173913: gap of unknown length
173914 174940: contig of 1027 bp in length
174941 175020: gap of unknown length
175021 175732: contig of 712 bp in length
175733 175813: gap of unknown length
175813 176895: contig of 1083 bp in length
176896 176975: gap of unknown length
176976 177582: contig of 607 bp in length
177583 177662: gap of unknown length
177663 178111: contig of 449 bp in length
178112 178191: gap of unknown length
178192 178900: contig of 709 bp in length
178901 178980: gap of unknown length

Query Match 34.6%; Score 18; DB 2; Length 198282;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 CATGCTCTGCTGCTGC 38

Db 93843 CATCGCTCTCGTCGCTGC 93860

RESULT 39			
AC007415/c			
LOCUS	198282 bp	DNA	linear
DEFINITION	Drosophila melanogaster chromosome 2 clone BACR05A24 (D603) RPCT-99	HTG 17-MAR-2000	

ACCESSION	AC007415
VERSION	AC007415.11
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	<i>Drosophila melanogaster</i> (fruit fly)
ORGANISM	<i>Drosophila melanogaster</i>

REFERENCE
AUTHORS
(bases 1 to 198282)
Celinkert,S.E., Agbayan,A., Arcaina,T.T., Baxter,E., Blazet,R.G.,
Buenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Paciel,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sechi,H., Svartkas,R.R., Van,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
TITLE
JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 198282)
Celinkert,S.E., Agbayan,A., Arcaina,T.T., Baxter,E., Blazet,R.G.

*	1	926:	contig of 926 bp in length
*	927	1006:	gap of unknown length
*	1007	1301:	contig of 285 bp in length
*	1302	1381:	gap of unknown length
*	1382	1860:	contig of 479 bp in length
*	1861	1940:	gap of unknown length
*	1941	2403:	contig of 463 bp in length
*	2404	2483:	gap of unknown length
*	2484	2895:	contig of 412 bp in length
*	2896	2975:	gap of unknown length
*	2976	3478:	contig of 503 bp in length
*	3479	3558:	gap of unknown length
*	3559	4034:	contig of 476 bp in length
*	4035	4114:	gap of unknown length
*	4115	4788:	contig of 674 bp in length
*	4789	4868:	gap of unknown length
*	4869	5549:	contig of 681 bp in length

5550	5629:	gap of unknown length
5630	6443:	contig of 814 bp in length
6444	6533:	gap of unknown length
6524	7697:	contig of 1174 bp in length
7698	7777:	gap of unknown length
7778	9243:	contig of 1466 bp in length
9244	9323:	gap of unknown length
9324	12162:	contig of 2839 bp in length
12163	12242:	gap of unknown length
12263	14263:	contig of 2021 bp in length
14264	14143:	gap of unknown length
14344	16143:	contig of 1761 bp in length
16105	16184:	gap of unknown length
16185	18228:	contig of 2044 bp in length
18229	18308:	gap of unknown length
18309	20874:	contig of 2566 bp in length
20875	20954:	gap of unknown length
20955	24951:	contig of 3997 bp in length
24952	25031:	gap of unknown length
25032	28493:	contig of 3462 bp in length
28494	28573:	gap of unknown length
28574	31026:	contig of 2453 bp in length
31027	31106:	gap of unknown length
31107	35546:	contig of 4440 bp in length
35547	35626:	gap of unknown length
35627	40849:	contig of 5223 bp in length
40850	40929:	gap of unknown length
40930	48761:	contig of 7832 bp in length
48762	48841:	gap of unknown length
48842	52556:	contig of 3815 bp in length
52557	52736:	gap of unknown length
52737	58188:	contig of 5452 bp in length
58189	58266:	gap of unknown length
58269	62896:	contig of 4628 bp in length
62897	62976:	gap of unknown length
62977	72175:	contig of 9199 bp in length
72176	72255:	gap of unknown length
72256	86109:	contig of 13654 bp in length
86110	86189:	gap of unknown length
86190	94772:	contig of 8533 bp in length
94773	94852:	gap of unknown length
94853	112088:	contig of 17236 bp in length
112089	112168:	gap of unknown length
112169	128075:	contig of 15907 bp in length
128076	128155:	gap of unknown length
128156	146641:	contig of 18486 bp in length
146642	146721:	gap of unknown length
146722	147805:	contig of 1084 bp in length
147806	147885:	gap of unknown length
147886	148190:	contig of 305 bp in length
148191	148270:	gap of unknown length
148271	149354:	contig of 984 bp in length
149355	149925:	gap of unknown length
149926	150046:	contig of 712 bp in length
150047	150126:	gap of unknown length
150127	151028:	contig of 902 bp in length
151029	151108:	gap of unknown length
151109	151116:	contig of 608 bp in length
151117	151796:	gap of unknown length
151797	152435:	contig of 659 bp in length
152436	152535:	gap of unknown length
152536	153204:	contig of 669 bp in length
153205	153284:	gap of unknown length
153285	153938:	contig of 654 bp in length
153939	154019:	gap of unknown length
154020	154280:	contig of 262 bp in length
154281	154360:	gap of unknown length
154361	155656:	contig of 1296 bp in length
155657	155736:	gap of unknown length
155737	155328:	contig of 592 bp in length
155329	156408:	gap of unknown length
156409	157266:	contig of 858 bp in length
157267	157346:	gap of unknown length

```

* 157347 157827: contig of 481 bp in length
* 157828 157907: gap of unknown length
* 157908 158993: contig of 1086 bp in length
* 158994 159073: gap of unknown length
* 159074 159316: contig of 243 bp in length
* 159317 159396: gap of unknown length
* 159397 159756: contig of 360 bp in length
* 159757 159836: gap of unknown length
* 159837 160697: contig of 861 bp in length
* 160698 160777: gap of unknown length
* 160778 161496: contig of 719 bp in length
* 161497 161576: gap of unknown length
* 161577 162155: contig of 579 bp in length
* 162156 162235: gap of unknown length
* 162236 162542: contig of 307 bp in length
* 162543 162622: gap of unknown length
* 162623 162964: contig of 342 bp in length
* 162965 163044: gap of unknown length
* 163045 163544: contig of 500 bp in length
* 163545 163625: gap of unknown length
* 163626 165092: contig of 1468 bp in length
* 165093 165172: gap of unknown length
* 165173 165866: contig of 694 bp in length
* 165867 165946: gap of unknown length
* 165947 166183: contig of 237 bp in length
* 166184 166263: gap of unknown length
* 166264 167263: contig of 1000 bp in length
* 167264 167343: gap of unknown length
* 167344 167883: contig of 540 bp in length
* 167884 167963: gap of unknown length
* 167964 168338: contig of 375 bp in length
* 168339 168418: gap of unknown length
* 168419 168719: contig of 301 bp in length
* 168720 168799: gap of unknown length
* 168800 169480: contig of 681 bp in length
* 169481 169560: gap of unknown length
* 169561 170057: contig of 507 bp in length
* 170058 170147: gap of unknown length
* 170148 170904: contig of 757 bp in length
* 170905 170984: gap of unknown length
* 170985 171769: contig of 785 bp in length
* 171770 171849: gap of unknown length
* 171850 172398: contig of 549 bp in length
* 172399 172478: gap of unknown length
* 172479 173200: contig of 722 bp in length
* 173201 173280: gap of unknown length
* 173281 173833: contig of 553 bp in length
* 173834 173913: gap of unknown length
* 173914 174940: contig of 1027 bp in length
* 174941 175020: gap of unknown length
* 175021 175732: contig of 712 bp in length
* 175733 175812: gap of unknown length
* 175813 176895: contig of 1083 bp in length
* 176896 176975: gap of unknown length
* 176976 177582: contig of 607 bp in length
* 177583 177662: gap of unknown length
* 177663 178111: contig of 449 bp in length
* 178112 178191: gap of unknown length
* 178192 178900: contig of 709 bp in length
* 178901 178980: gap of unknown length

```

Query Match 34.6%; Score 18; DB 2; Length 198282;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 CATCGCTCTGCTGCTGC 38
 Db 146075 CATCGCTCTGCTGCTGC 146058

RESULT 40
 AC013831
 LOCUS AC013831 201451 bp DNA linear HTG 15-MAR-2002

```

DEFINITION
Drosophila melanogaster chromosome 2 clone BAC112A16 (D1172)
RP11-98 12.A.16 map 40D-40D strain Y; cn bw sp, *** SEQUENCING IN
PROCESS ***; 136 unordered pieces.
ACCESSION
AC013831
VERSION
AC013831.13 GI:19482321
KEYWORDS
HTG, HTGS, PHASE1
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 201451)
REFERENCE
Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svitskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 201451)
REFERENCE
Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svitskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Direct Submission
Submitted (15-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
3 (bases 1 to 201451)
REFERENCE
Celiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svitskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Direct Submission
Submitted (15-MAR-2002) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
NOTE: This is a 'working draft' sequence. It currently
consists of 136 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
622: contig of 622 bp in length
623
702: gap of unknown length
703
1319: contig of 617 bp in length
1320
1399: gap of unknown length
1400
2019: contig of 619 bp in length
2099
2696: gap of unknown length
2697
2776: gap of unknown length
2777
3421: contig of 645 bp in length
3422
4078: contig of 577 bp in length
4079
4158: gap of unknown length
4159
4671: contig of 513 bp in length

```

* 4672 4751: gap of unknown length
* 4752 5517: contig of 766 bp in length
* 5518 5597: gap of unknown length
* 5598 6243: contig of 646 bp in length
* 6244 6323: gap of unknown length
* 6324 6965: contig of 642 bp in length
* 6966 7045: gap of unknown length
* 7046 7659: contig of 614 bp in length
* 7660 7739: gap of unknown length
* 7740 8525: contig of 786 bp in length
* 8526 8605: gap of unknown length
* 8606 9300: contig of 695 bp in length
* 9301 9380: gap of unknown length
* 9381 10413: contig of 1033 bp in length
* 10414 10493: gap of unknown length
* 10494 11059: contig of 566 bp in length
* 11060 11139: gap of unknown length
* 11140 12030: contig of 891 bp in length
* 12031 12110: gap of unknown length
* 12111 12820: contig of 710 bp in length
* 12821 12900: gap of unknown length
* 12901 13575: contig of 675 bp in length
* 13576 13655: gap of unknown length
* 13656 14388: contig of 733 bp in length
* 14389 14468: gap of unknown length
* 14469 15443: contig of 975 bp in length
* 15444 15523: gap of unknown length
* 15524 16504: contig of 981 bp in length
* 16505 16584: gap of unknown length
* 16585 18119: contig of 1535 bp in length
* 18120 18199: gap of unknown length
* 18200 19453: contig of 1254 bp in length
* 19454 19533: gap of unknown length
* 19534 20795: contig of 1262 bp in length
* 20796 20875: gap of unknown length
* 20876 22136: contig of 1261 bp in length
* 22137 22216: gap of unknown length
* 22217 23747: contig of 1531 bp in length
* 23748 23827: gap of unknown length
* 23828 25601: contig of 1774 bp in length
* 25602 25681: gap of unknown length
* 25682 28215: contig of 2534 bp in length
* 28216 28295: gap of unknown length
* 28296 29804: contig of 1509 bp in length
* 29805 29884: gap of unknown length
* 29885 31373: contig of 1489 bp in length
* 31374 31453: gap of unknown length
* 31454 33841: contig of 2388 bp in length
* 33842 33921: gap of unknown length
* 33922 37415: contig of 3494 bp in length
* 37416 37495: gap of unknown length
* 37496 40096: contig of 2601 bp in length
* 40097 40176: gap of unknown length
* 40177 44443: contig of 4267 bp in length
* 44444 44523: gap of unknown length
* 44524 47877: contig of 3354 bp in length
* 47878 47957: gap of unknown length
* 47958 52119: contig of 4162 bp in length
* 52120 52199: gap of unknown length
* 52200 56290: contig of 4091 bp in length
* 56291 56370: gap of unknown length
* 56371 62684: contig of 6314 bp in length
* 62685 62764: gap of unknown length
* 62765 71961: contig of 9197 bp in length
* 71962 72041: gap of unknown length
* 72042 79797: contig of 7756 bp in length
* 79798 79877: gap of unknown length
* 79878 101310: contig of 21433 bp in length
* 101311 101390: gap of unknown length
* 101391 109584: contig of 8194 bp in length
* 109585 109664: gap of unknown length
* 109665 121016: contig of 13352 bp in length
* 121017 121096: gap of unknown length

* 121097 137479: contig of 16383 bp in length
* 137480 137559: gap of unknown length
* 137560 138592: contig of 1033 bp in length
* 138593 138672: gap of unknown length
* 138673 139477: contig of 805 bp in length
* 139478 139557: gap of unknown length
* 139558 140520: contig of 963 bp in length
* 140521 140600: gap of unknown length
* 140601 141218: contig of 618 bp in length
* 141219 141298: gap of unknown length
* 141299 141969: contig of 671 bp in length
* 141970 142049: gap of unknown length
* 142050 142463: contig of 414 bp in length
* 142464 142543: gap of unknown length
* 142544 143220: contig of 677 bp in length
* 143221 143300: gap of unknown length
* 143301 143706: contig of 406 bp in length
* 143707 143786: gap of unknown length
* 143787 144114: contig of 328 bp in length
* 144115 144194: gap of unknown length
* 144195 144798: contig of 604 bp in length
* 144799 144878: gap of unknown length
* 144879 145387: contig of 509 bp in length
* 145388 145467: gap of unknown length
* 145468 146605: contig of 1138 bp in length
* 146606 146685: gap of unknown length
* 146686 146934: contig of 249 bp in length
* 146935 147014: gap of unknown length
* 147015 147251: contig of 237 bp in length
* 147252 147331: gap of unknown length
* 147332 148015: contig of 684 bp in length
* 148016 148095: gap of unknown length
* 148096 148714: contig of 619 bp in length
* 148715 148794: gap of unknown length
* 148795 149416: contig of 622 bp in length
* 149417 149496: gap of unknown length
* 149497 150142: contig of 646 bp in length
* 150143 150222: gap of unknown length
* 150223 150916: contig of 694 bp in length
* 150917 150996: gap of unknown length
* 150997 151559: contig of 563 bp in length
* 151560 151639: gap of unknown length
* 151640 152228: contig of 589 bp in length
* 152229 152308: gap of unknown length
* 152309 152918: contig of 610 bp in length
* 152919 152998: gap of unknown length
* 152999 153639: contig of 641 bp in length
* 153640 153719: gap of unknown length
* 153720 154031: contig of 312 bp in length
* 154032 154111: gap of unknown length
* 154112 154581: contig of 470 bp in length
* 154582 154661: gap of unknown length
* 154662 154873: contig of 212 bp in length

Query Match 34.6% Score 18: DB 2: Length 201451:
Best Local Similarity 100.0%: Pred. No. 36;
Matches 18: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 CATCGCTCTCGTCGCTGC 38
DB 106212 CATCGCTCTCGTCGCTGC 106229

Search completed: November 13, 2003, 10:26:37
Job time : 2042 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 08:53:26 : Search time 217 Seconds
(without alignments)
646.870 Million cell updates/sec

Title: US-10-081-935-1

Perfect score: 52
Sequence: 1 atgcagctctctccacgctct.....cgctgcgcgcctccagctg 52

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

N_Geneseq_190jun03.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19	36.5	3416	21	AA552247
C 2	17	32.7	1392	25	AA050552
C 3	17	32.7	2684	23	ABU19972
C 4	17	32.7	4310	23	ABU15450
5	16	30.8	276	21	AAA31959
6	16	30.8	357	21	AAA31675
7	16	30.8	438	21	AAA31705
C 8	16	30.8	598	23	AA584162

C 9	16	30.8	638	21	AA035333	Arabidopsis thaliana DNA encoding novel
10	16	30.8	1582	23	AA569014	Human leucine-rich
11	16	30.8	1653	24	AA028124	Human secreted pro
12	16	30.8	1872	22	AA076853	Human Lp220 secret
13	16	30.8	2369	24	AA038695	Human NOV44b codin
14	16	30.8	2451	24	ABV99421	Human NOV44c codin
15	16	30.8	2563	24	ABV99422	Human NOV44c codin
16	16	30.8	2569	25	AA099609	MDMT related human
17	16	30.8	3201	24	ABV99430	Human NOV44a codin
C 18	16	30.8	3947	21	AA016203	Human prostate can
C 19	16	30.8	3953	22	AA034839	Human colon cancer
C 20	16	30.8	6507	23	ABU15493	Drosophila melanog
C 21	16	30.8	7434	23	AA584163	Drosophila melanog
C 22	16	30.8	15435	23	ABU14186	Drosophila melanog
C 23	16	30.8	47999	25	AA052898	Human tucetly homol
C 24	16	30.8	59967	23	ABU15492	Drosophila melanog
C 25	16	30.8	349980	24	ABO81844	Bifidobacterium 10
C 26	15	28.8	214	24	ABL74833	Corn tassal-derive
C 27	15	28.8	264	25	ABX33704	Human GDP-mannose
C 28	15	28.8	335	25	ABX17641	S2 subtracction lib
C 29	15	28.8	336	25	ABX19200	Human GDP-mannose
C 30	15	28.8	357	22	AA531315	Human cDNA encodin
C 31	15	28.8	357	23	ABK43452	DNA encoding novel
C 32	15	28.8	357	23	ABK43452	Human polynucleoti
C 33	15	28.8	416	23	AA576408	DNA encoding novel
C 34	15	28.8	423	20	AA219396	M. tuberculosis an
C 35	15	28.8	423	20	AA219396	M. tuberculosis re
C 36	15	28.8	448	21	ABO62561	Mycobacterium tube
C 37	15	28.8	459	21	ABO62561	Human secreted pro
C 38	15	28.8	471	21	AA020201	Human foetal liver
C 39	15	28.8	481	22	ABX58638	Probe #6099 for ge
C 40	15	28.8	481	22	ABX58638	Human bone marrow
C 41	15	28.8	481	22	AA116089	Human brain expres
C 42	15	28.8	481	22	AA116089	Probe #6022 used t
C 43	15	28.8	481	22	AA116089	Probe #6998 used t
C 44	15	28.8	481	23	ABX32191	Human liver single
C 45	15	28.8	481	23	ABX32191	Human genome-deriv
C 46	15	28.8	481	23	ABX32191	Human immune-deriv
C 47	15	28.8	619	22	AA080662	Human cDNA clone (
C 48	15	28.8	667	22	AA080662	Aspergillus oryzae
C 49	15	28.8	695	21	AA012705	Human cDNA clone (
C 50	15	28.8	753	22	AA012705	M. capsulatus gene
C 51	15	28.8	758	24	ABO90246	Human polynucleoti
C 52	15	28.8	1197	22	AA158291	Human polynucleoti
C 53	15	28.8	1340	21	AA000690	Human H2O1ase pr
C 54	15	28.8	1358	22	AA541256	cDNA encoding nove
C 55	15	28.8	1366	23	ABV23063	Human prostate exp
C 56	15	28.8	1386	23	ABV28899	Human prostate exp
C 57	15	28.8	1404	25	AA050288	Breast cancer asso
C 58	15	28.8	1431	24	AB070729	Breast cancer-asso
C 59	15	28.8	1435	25	ABX72219	Human NOVX polynuc
C 60	15	28.8	1446	22	AA046450	Pentacillin chryso
C 61	15	28.8	1473	24	ABO60922	Human h1pocampal
C 62	15	28.8	1502	21	AA033695	Arabidopsis thaliana
C 63	15	28.8	1507	21	AA033695	Arabidopsis thaliana
C 64	15	28.8	1554	25	AB020720	Aspergillus fumiga
C 65	15	28.8	1750	24	AB067741	Human intercellula
C 66	15	28.8	1778	25	AB020122	Aspergillus fumiga
C 67	15	28.8	1878	23	AA120539	Drosophila melanog
C 68	15	28.8	2000	24	AB215173	Drosophila thalia
C 69	15	28.8	2134	23	ABU02831	Drosophila melanog
C 70	15	28.8	2304	23	ABU17476	Drosophila melanog
C 71	15	28.8	2454	23	AA551556	Pseudomonas aerugi
C 72	15	28.8	2520	24	AB084687	Human HCCA2 encod
C 73	15	28.8	2556	22	AA017696	Human cDNA sequenc
C 74	15	28.8	2617	22	AA016680	Human secreted pro
C 75	15	28.8	2672	21	AA016680	Human HCCA1 encod
C 76	15	28.8	2702	24	ABV99189	Lung cancer relat
C 77	15	28.8	2790	24	ABU65208	Ovary cancer relat
C 78	15	28.8	2790	24	ABU65208	Drosophila melanog
C 79	15	28.8	3103	23	ABU08033	Aspergillus fumiga
C 80	15	28.8	3140	25	ABU17712	cDNA sequence #197
C 81	15	28.8	3323	24	ABK35806	

c 958	13	25.0	3189	23	AA551474	Pseudomonas aerugi
959	13	25.0	3220	22	AA534856	CDNA encoding nove
c 960	13	25.0	3239	23	ABU13684	Drosophila melanog
c 961	13	25.0	3252	23	AA585001	DNA encoding novel
c 962	13	25.0	3260	23	ABU02197	Drosophila melanog
c 963	13	25.0	3266	23	ABU16043	Drosophila melanog
c 964	13	25.0	3276	22	AAK52285	Human polynucleoti
c 965	13	25.0	3283	24	AB199253	Mouse ischaemic co
c 966	13	25.0	3285	22	AAH54777	S. epidermidis gen
c 967	13	25.0	3300	23	ABU29540	Drosophila melanog
c 968	13	25.0	3301	21	AA251259	Human RNA-associat
c 969	13	25.0	3305	23	ABU11137	Drosophila melanog
c 970	13	25.0	3307	23	ABU19010	S. epidermidis gen
971	13	25.0	3314	22	AAH54563	S. epidermidis gen
c 972	13	25.0	3333	24	AB190281	Human polynucleoti
c 973	13	25.0	3335	24	AB211773	Human polynucleoti
c 974	13	25.0	3351	20	AAZ31552	S. rochei strain E
c 975	13	25.0	3399	23	ABU09972	Drosophila melanog
c 976	13	25.0	3417	23	ABU29536	Drosophila melanog
977	13	25.0	3440	23	ABV21315	Human prostate exp
c 978	13	25.0	3440	23	ABV22159	Human prostate exp
c 979	13	25.0	3440	23	ABV27134	Human prostate exp
c 980	13	25.0	3440	23	ABU27998	Human prostate exp
c 981	13	25.0	3443	23	ABU12680	Drosophila melanog
c 982	13	25.0	3446	24	ABZ11772	Human polynucleoti
c 983	13	25.0	3467	22	AAFS9301	Aspergillus niger
c 984	13	25.0	3467	23	ABA92733	Aspergillus niger
c 985	13	25.0	3470	25	ABT19514	Aspergillus fumiga
c 986	13	25.0	3477	23	ABU21729	Drosophila melanog
c 987	13	25.0	3518	22	AAK52083	Human polynucleoti
c 988	13	25.0	3518	20	AAK38287	Herpes simplex vir
c 989	13	25.0	3540	23	ABU07654	Drosophila melanog
c 990	13	25.0	3556	22	AA159883	Human polynucleoti
c 991	13	25.0	3570	24	AB278220	A. niger kex gene.
c 992	13	25.0	3573	23	ABU05120	Drosophila melanog
c 993	13	25.0	3579	23	ABV21968	Human prostate exp
994	13	25.0	3585	21	AA97456	Human RalGDS (hral
c 995	13	25.0	3598	22	AAH54124	S. epidermidis gen
c 996	13	25.0	3601	24	ABK83884	Human CDNA differe
c 997	13	25.0	3608	25	ABT17648	Aspergillus fumiga
c 998	13	25.0	3632	22	AA005133	Human secreted pro
c 999	13	25.0	3668	23	ABU23328	Drosophila melanog
c1000	13	25.0	3682	23	ABU02641	Drosophila melanog

ALIGNMENTS

```

RESULT 1
AA252247/c
ID AA252247 standard; DNA; 3416 BP.
XX
AC AA252247;
XX
DT 18-JUL-2000 (first entry)
XX
DE Maize glycine-rich protein 3 partial genomic clone zmGRP3.
XX
KW Maize; glycine-rich protein 3; GRP3; regulatory element; zmGRP3;
KM root specific gene expression; root abundant gene; monocotyledon;
XX pathogen resistance; pest; herbicide; growth rate; ds.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT 1..1780 /*tag= a
FT TATA_signal 1712..1719 /*tag= b
FT CDS 1781..2560 /*tag= c
FT /*product= "GRP3"
FT /*note= "Glycine-rich protein 3"

```

```

FT misc_signal 2561..3416
FT /*tag= d
FT /*label= 3'_regulatory_element
FT polyA_signal 2692..2696
FT /*tag= e
PN WO200015662-A1.
PD 23-MAR-2000.
PF 10-SEP-1999; 99WO-EP06692.
PR 11-SEP-1998; 98EP-0117251.
PA (AGRI-) AGRIC TECHNOLOGY & GENETICS GMBH.
XX
XX Feix G, Muller D;
XX
XX WPI; 2000-271382/23.
XX
XX Nucleic acid sequence is used for cloning and expressing a root
XX specific or root abundant gene in a plant -
XX
XX Claim 4; Page 55-56; 60pp; English.
XX
XX The present sequence is a partial genomic clone zmGRP3 encoding
XX maize glycine-rich protein 3 (GRP3). This sequence comprises 5' and
XX 3' regulatory elements useful for cloning and
XX expressing root specific or root abundant genes in plants, especially
XX monocots which provide high expression efficiency and high
XX tissue specificity. Root preferred gene expression provides several
XX advantages to plants e.g. resistance to pathogens, pests, herbicides and
XX adverse weather conditions, modification of growth rate and alteration of
XX root tissue function. This sequence also provides a means of isolating
XX related regulatory sequences of other plant species which confer root
XX specificity to genes of interest operably linked to them.
XX
XX Sequence 3416 BP; 755 A; 841 C; 950 G; 870 T; 0 other;
XX
XX
XX Query Match 36.5%; Score 19; DB 21; Length 3416;
XX Best Local Similarity 100.0%; Pred. No. 0.62;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 30 CGTGGCTGCCGCGCTGCC 48
XX
XX Db 2429 CGTGGCTGCCGCGCTGCC 2411
XX
XX
XX RESULT 2
XX AAD50522/c
XX ID AAD50522 standard; DNA; 1392 BP.
XX
XX AC AAD50522;
XX
XX DT 24-MAR-2003 (first entry)
XX
XX DE Mycobacterium avium AST4 (mav_144) glycosyl sulfoltransferase DNA.
XX
XX KM Mycobacterial sulphation pathway; immune response; infection; vaccine;
XX gene therapy; glycosyl sulfoltransferase; enzyme; AST4 protein; gene; ds.
XX
XX OS Mycobacterium avium.
XX
XX FH Key Location/Qualifiers
XX FT 1..1392 /*tag= a
XX FT CDS /*product= "AST4 protein"
XX
XX WO200286067-A2.
XX
XX 31-OCT-2002.
XX
XX 19-APR-2002; 2002WO-US12374.
XX

```

```
XX 20-APR-2001; 2001US-285394P.
PR 26-OCT-2001; 2001US-345953P.
XX (REGC ) UNIV CALIFORNIA.
XX Bertozi C, Williams SJ, Mougous J;
XX WPI: 2001-093116/08.
XX P-PSDB; AAE32776.
DR
XX Novel mycobacterial sulfation pathway polypeptide useful in vitro
PT cell-free assay for identifying agent that reduces the activity of the
PT polypeptide -
XX
XX Disclosure; Page 114-116; 141pp; English.
PS
XX The present invention relates to mycobacterial sulphation pathway enzymes
CC and polynucleotides encoding such proteins. Sequences of the invention
CC are useful in vitro cell-free assay for identifying agents that reduce
CC the activity of the proteins. They are useful for increasing an immune
CC response to pathogenic mycobacterium in a host. They are also useful for
CC treating mycobacterial infection in mammals. Sulphation pathway proteins
CC are useful for reducing the virability and virulence of a mycobacterium.
CC They are used as vaccines and in gene therapy. The present sequence is
CC Mycobacterium avium AST4 (may_144) glycosyl sulfotransferase DNA.
XX
SQ Sequence 1392 BP; 215 A; 485 C; 481 G; 211 T; 0 other;
QY
Query Match 32.7%; Score 17; DB 25; Length 1392;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 29 TCGTCGTCGCCGCCCTC 45
290 TCGTCGTCGCCGCCCTC 274
RESULT 3
ABL19972/C
ID ABL19972 standard; DNA; 2684 BP.
XX
XX ABL19972;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 11389.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEXE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 11389; 21pp + Sequence Listing; English.
```

```
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2684 BP; 807 A; 534 C; 541 G; 802 T; 0 other;
QY
Query Match 32.7%; Score 17; DB 23; Length 2684;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 7 TTCTTCACGCTTCTCAT 23
2101 TTCTTCACGCTTCTCAT 2085
RESULT 4
ABL15450/C
ID ABL15450 standard; cDNA; 4310 BP.
XX
XX ABL15450;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 40832.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEXE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX
XX P-PSDB; ABB71347.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 40832; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
```

SQ Sequence 4310 BP; 1181 A; 938 C; 946 G; 1245 T; 0 other;
 Query Match 32.7%; Score 17; DB 23; Length 4310;
 Best Local Similarity 100.0%; Pred. No. 7.5;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 TTCTCTACGCTCTCAT 23
 |||||
 DB 848 TTCTCTACGCTCTCAT 832
 |||||
 RESULT 5
 AAA31959
 ID AAA31959 standard; DNA; 276 BP.
 XX
 AC AAA31959;
 XX
 DT 05-JUL-2000 (first entry)
 XX
 DE Plant microsatellite marker #920.
 XX
 KM Plant microsatellite sequence; core repeat sequence; detection; probe;
 KM DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KM variety identification; genetic variability evaluation; primer; ss.
 XX
 OS Eucalyptus grandis.
 XX
 PN MO9967421-A1.
 XX
 PD 29-DEC-1999.
 XX
 PF 25-JUN-1999; 99WO-NZ00092.
 XX
 PR 25-JUN-1998; 98US-0105307.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Havukkala IJ, Bloksberg LN, Glenn M;
 PI
 DR WPI; 2000-116958/10.
 XX
 PT New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -
 XX
 PS Claim 1; Page 347; 392pp; English.
 XX
 CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.
 XX
 SO Sequence 276 BP; 60 A; 82 C; 88 G; 46 T; 0 other;
 Query Match 30.8%; Score 16; DB 21; Length 276;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 21 CATGCTCTCTGCTGCT 36
 |||||
 DB 157 CATGCTCTCTGCTGCT 172
 |||||
 RESULT 6
 AAA31675

ID AAA31675 standard; DNA; 357 BP.
 XX
 AC AAA31675;
 XX
 DT 05-JUL-2000 (first entry)
 XX
 DE Plant microsatellite marker #636.
 XX
 KM Plant microsatellite sequence; core repeat sequence; detection; probe;
 KM DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KM variety identification; genetic variability evaluation; primer; ss.
 XX
 OS Eucalyptus grandis.
 XX
 PN MO9967421-A1.
 XX
 PD 29-DEC-1999.
 XX
 PF 25-JUN-1999; 99WO-NZ00092.
 XX
 PR 25-JUN-1998; 98US-0105307.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Havukkala IJ, Bloksberg LN, Glenn M;
 PI
 DR WPI; 2000-116958/10.
 XX
 PT New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -
 XX
 PS Claim 1; Page 264; 392pp; English.
 XX
 CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.
 XX
 SO Sequence 357 BP; 78 A; 92 C; 109 G; 78 T; 0 other;
 Query Match 30.8%; Score 16; DB 21; Length 357;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 21 CATGCTCTCTGCTGCT 36
 |||||
 DB 154 CATGCTCTCTGCTGCT 169
 |||||
 RESULT 7
 AAA31705
 ID AAA31705 standard; DNA; 438 BP.
 XX
 AC AAA31705;
 XX
 DT 05-JUL-2000 (first entry)
 XX
 DE Plant microsatellite marker #666.
 XX
 KM Plant microsatellite sequence; core repeat sequence; detection; probe;
 KM DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KM variety identification; genetic variability evaluation; primer; ss.
 XX
 OS Eucalyptus grandis.

```
XX PN WO9967421-A1.
XX
XX PD 29-DEC-1999.
XX
XX PF 25-JUN-1999; 99WO-NZ00092.
XX
XX PR 25-JUN-1999; 98US-0105307.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX PI Havukkala IJ, Bloksberg LN, Glenn M;
XX
XX DR WPI; 2000-116958/10.
XX
XX PT New plant microsatellite markers and associated flanking species for
XX the detection of polymorphic genetic markers -
XX
XX PS Claim 1; Page 273; 392pp; English.
XX
CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences AAA32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
SQ Sequence 438 BP; 100 A; 107 C; 125 G; 106 T; 0 other;
Query Match 30.8%; Score 16; DB 21; Length 438;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 CATCGCTCTCGTGGCT 36
DB 149 CATCGCTCTCGTGGCT 164
RESULT 8
AAS84162/C
ID AAS84162 standard; cDNA; 598 BP.
XX
XX AC AAS84162;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #19966.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Drymanac RT, Liu C, Tang YT;
XX
```

```
DR WPI; 2001-639362/73.
DR P-PSDB; ABG19975.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX PS Claim 1; SEQ ID NO 19966; 103pp; English.
XX
XX CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 598 BP; 144 A; 166 C; 171 G; 117 T; 0 other;
Query Match 30.8%; Score 16; DB 23; Length 598;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 CTCATCGCTCTCGTCG 34
DB 63 CTCATCGCTCTCGTCG 48
RESULT 9
AAC35333/C
ID AAC35333 standard; DNA; 638 BP.
XX
XX AC AAC35333;
XX
XX DT 17-OCT-2000 (first entry)
XX
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9808.
XX
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX OS Arabidopsis thaliana.
XX
XX PN EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX PR 25-FEB-1999; 99US-0121825.
XX
XX PR 05-MAR-1999; 99US-0123180.
XX
XX PR 09-MAR-1999; 99US-0123548.
XX
XX PR 23-MAR-1999; 99US-0125788.
XX
XX PR 25-MAR-1999; 99US-0126264.
XX
XX PR 29-MAR-1999; 99US-0126785.
XX
XX PR 01-APR-1999; 99US-0127462.
XX
XX PR 06-APR-1999; 99US-0128234.
```


PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142280.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.

```
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

```
Query Match 30.8%; Score 16; DB 21; Length 638;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGAGTCTCTCAGC 16
Db 609 ATCGAGTCTCTCAGC 594
```

```
RESULT 10
AAS69014
ID AAS69014 standard; cDNA; 1582 BP.
XX
AC AAS69014;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #4818.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB; ABG04827.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 4818; 103pp; English.
```

```
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
```

```
SQ Sequence 1582 BP; 216 A; 603 C; 483 G; 279 T; 1 other;

Query Match 30.8%; Score 16; DB 23; Length 1582;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 CGCTCGCGGCTCGCC 48
Db 1380 CGCTCGCGGCTCGCC 1395
```

```
RESULT 11
AAD28124
ID AAD28124 standard; DNA; 1653 BP.
XX
XX
XX AAD28124;
XX
XX
DT 22-APR-2002 (first entry)
XX
DE Human leucine-rich repeat-8 (ZLR8) DNA #2.
XX
KW Human; leucine-rich repeat-8; ZLR8; cytosstatic; gene therapy; leukaemia;
KW endometrial adenocarcinoma; renal cell; colon; prostate; retinoblastoma;
KW cell growth disorder; glioblastoma; neuroblastoma; kidney tumour; lung;
KW mammary gland; germ cell; bladder; oesophagus; pancreas; animal feed;
KW rhabdomyosarcoma; genitourinary tract; chromosome 11q13; ds.
XX
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX CDS 1..1653
XX FT /*tag= a
XX FT /product= "ZLR8 protein"
XX FT /note= "CDS does not include stop codon"
XX FT /partial
XX FT sig_peptide 1..54
XX FT /*tag= b
XX FT mat_peptide 55..1653
XX FT /*tag= c
XX FT /product= "Mature ZLR8 protein"
XX
XX
XX WO200202604-A2.
XX
XX
XX 10-JAN-2002.
XX
XX
XX 02-JUL-2001; 2001WO-US20999.
XX
XX 30-JUN-2000; 2000US-215446P.
XX
XX (ZYMO ) ZYMOGENETICS INC.
XX
XX
```

```
XX Thayer EC, Sheppard PO, Presnell SR;
PI WPI: 2002-154725/20.
XX P-PSDB; AAE17484.
DR
XX New leucine-rich repeat proteins and polynucleotides, useful for
PT diagnosing and treating disorders related to abnormal cell growth e.g.
PT retinoblastoma, renal cell, endometrial adenocarcinoma, leukemia,
PT kidney and lung tumors -
XX
XX Claim 19: Page 64-67; 82pp; English.
XX
XX The invention relates to an isolated polypeptide comprising leucine-rich
CC repeat proteins Zlir7, Zlir8, Zlir9 and spliced variants of Zlir7, Zlir9,
CC Zlir7, Zlir8, and Zlir9 proteins are useful in directing the secretion of
CC proteins of interest from a host cell and to monitor the secretion of
CC proteins in general from cells and tissues. The Zlir DNA and proteins
CC are useful in diagnosing and treating disorders related to abnormal cell
CC growth, including retinoblastoma, renal cell adenocarcinoma, endometrial
CC adenocarcinoma, glioblastoma, neuroblastoma, B-cell lymphocytic leukaemia,
CC kidney tumors, germ cell tumors, lung large cell carcinoma, mammary,
CC colon adenocarcinoma, genitourinary tract transitional cell tumours,
CC colon adenocarcinoma, lung tumour, bladder tumour, oesophagus, pancreas
CC and prostate adenocarcinoma. Zlir protein is useful for identifying
CC agonists and antagonists of the polypeptide, for drug design, to screen
CC for cell metabolism affecting receptors, for analysis of cell phenotype,
CC and as animal feed supplement and cell culture components. Zlir DNA is
CC also useful in gene therapy. The present sequence is human Zlir8 DNA.
CC
XX Zlir8 gene is located on chromosome 11q13.
XX
SQ Sequence 1653 BP; 223 A; 617 C; 526 G; 287 T; 0 other;
XX
Query Match 30.8%; Score 16; DB 24; Length 1653;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 33 CGCTCGCGGCTCGCC 48
DB 1496 CGCTCGCGGCTCGCC 1511
XX
RESULT 12
AAF76853
ID AAF76853 standard; cDNA; 1872 BP.
XX
XX AAF76853;
XX
DE 14-MAY-2001 (first entry)
XX
XX Human secreted protein cDNA #11.
XX
XX Human; secreted protein; immunomodulatory; antisclerotic;
XX immunosuppressive; anti-inflammatory; anti-HIV;
XX immunostimulant; cytostatic; cardiac; vascular; anti-angiogenic;
XX ophthalmological; neuroprotectant; nocitropic; anticonvulsant; vaccine;
XX antialzheimers; antiparkinsonian; antimicrobial; vulnertary; gene therapy;
XX immune disorder; hyperproliferative; cardiovascular; angiogenic;
XX neurological; infection; ss.
XX
XX Homo sapiens.
XX OS
XX WO200112776-A2.
XX PN
XX 22-FEB-2001.
XX PD
XX 15-AUG-2000; 2000WO-US22350.
XX PF
XX 16-AUG-1999; 99US-0148759.
XX PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX Shi Y, Young PE, Ebner R, Soppet DR, Ruben SM;
XX PI
```

```
XX WPI: 2001-244245/25.
DR P-PSDB; AAB70072.
XX
XX Nucleic acids encoding 18 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. Cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
XX Claim 1: Page 354-355; 380pp; English.
XX
XX The present sequence is one of 18 nucleic acid molecules encoding novel
CC human secreted proteins. The nucleic acids and proteins may be used in
CC the prevention, diagnosis and treatment of diseases including immune
CC disorders (e.g. multiple sclerosis, systemic lupus erythematosus and
CC human immunodeficiency virus (HIV) infections), hyperproliferative
CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC neovascularisation and diabetic retinopathy), neurological disorders
CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC infectious diseases and/or for promoting wound healing, regeneration
CC and/or chemotaxis. The nucleic acid molecules may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples. The polypeptides may also be
CC used as antigens in the production of antibodies and in assays to
CC identify modulators of protein expression and activity.
XX
SQ Sequence 1872 BP; 290 A; 659 C; 604 G; 319 T; 0 other;
XX
Query Match 30.8%; Score 16; DB 22; Length 1872;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 33 CGCTCGCGGCTCGCC 48
DB 1153 CGCTCGCGGCTCGCC 1168
XX
RESULT 13
AAD38695
ID AAD38695 standard; cDNA; 2369 BP.
XX
XX AAD38695;
XX
DE 23-SEP-2002 (first entry)
XX
XX Human LP220 secreted protein encoding cDNA.
XX
XX Human; secreted protein; atherosclerosis; Alzheimer's disease; LP220;
XX diabetic retinopathy; severe combined immunodeficiency; pancreatitis;
XX rheumatoid arthritis; colorectal adenoma; haemolytic anaemia; cancer;
XX reperfusion injury; arteriosclerosis; wound healing; transgenic animal;
XX gene therapy; neoplasm; transgenic; psoriasis; ischemia; carcinoma;
XX chromosome 11q13; gene; ss.
XX
XX Homo sapiens.
XX OS
XX Key Location/Qualifiers
XX FH CDS 167..2074
XX FT /tag= a
XX FT /product= "human LP220 secreted protein"
XX FT sig_peptide 167..214
XX FT /tag= b
XX FT mat_peptide 215..2071
XX FT /tag= c
XX FT /product= "mature human LP220 secreted protein"
XX
XX WO200226801-A2.
XX PN
XX 04-APR-2002.
XX PD
XX 14-SEP-2001; 2001WO-US26026.
XX PF
XX
```

```

PR 28-SEP-2000; 2000US-236088P.
XX (ELIL ) LILLY & CO ELI.
PA
XX Su EW, Wang H;
P1
XX WPI; 2002-471259/50.
DR P-PSDB; AAE23980.
XX
PT Novel proteins and polynucleotides of secreted proteins useful for
PT treating various diseases e.g. rheumatoid arthritis, cancer, psoriasis,
PT diabetic retinopathy, arteriosclerosis, ischemia or reperfusion injury
PT
XX
XX Claim 1; Page 124-127; 145pp; English.
PS
XX The invention relates to human secreted polypeptides designated LP095,
CC LP191, LP217, LP220, LP221, LP222, LP229, LP237 or LP238 and nucleic
CC acid molecules encoding such polypeptides. Novel secreted proteins of
CC the invention are used for treating diseases such as arteriosclerosis,
CC Alzheimer's disease, diabetic retinopathy, psoriasis, pancreatitis,
CC arteriosclerosis, rheumatoid arthritis, colorectal adenomas, severe
CC combined immunodeficiency, ischaemia, carcinomas, haemolytic anaemia,
CC reperfusion injury, neoplasms and cancer especially liver cancer. They
CC are also used for wound healing. Polynucleotides of the invention can
CC be used to generate transgenic animals or knock out animals, which in
CC turn, are useful in the development and screening of therapeutically
CC useful reagents for use in the treatment of diseases associated with
CC LP polypeptide associated activity. They are also used in gene therapy.
CC The present sequence is human LP220 secreted protein encoding cDNA.
CC LP220 gene is located on chromosome 11q13.
XX
SQ Sequence 2369 BP; 342 A; 855 C; 755 G; 417 T; 0 other;
Query Match 30.8%; Score 16; DB 24; Length 2369;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 33 CGCTGCGGCGCTCGCC 48
Db 1662 CGCTGCGGCGCTCGCC 1677
RESULT 14
ABV99421
ID ABV99421 standard; DNA; 2451 BP.
XX
AC ABV99421;
XX
DT 27-JAN-2003 (first entry)
XX
DE Human NOVA4b coding sequence.
XX
XX Human; anti-HIV; cytostatic; antidiabetic; antiashtmatic; cachexia; AIDS;
KW antiinflammatory; cardiac; haemostatic; neuroprotective; anorectic;
KW nootropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;
KW antifertility; cerebroprotective; gene therapy; NOXV; NOV; fertility;
KW metabolic disorder; diabetes; obesity; infectious disease; anorexia;
KW neurodegenerative disease; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; cardiovascular disorder;
KW bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;
KW metabolic syndrome X; wasting disorder; cell differentiation; gene;
KW cell proliferation; haematopoiesis; wound healing; angiogenesis; ds.
XX
XX Homo sapiens.
OS
XX WO200272771-A2.
PN
XX 19-SEP-2002.
PD
XX 08-MAR-2002; 2002WO-US07288.
PF
XX 08-MAR-2001; 2001US-274101P.
PR

```

```

PR 08-MAR-2001; 2001US-274194P.
PR 08-MAR-2001; 2001US-274281P.
PR 08-MAR-2001; 2001US-274322P.
PR 09-MAR-2001; 2001US-274849P.
PR 12-MAR-2001; 2001US-275235P.
PR 13-MAR-2001; 2001US-275578P.
PR 13-MAR-2001; 2001US-275579P.
PR 13-MAR-2001; 2001US-275601P.
PR 14-MAR-2001; 2001US-276000P.
PR 14-MAR-2001; 2001US-276776P.
PR 19-MAR-2001; 2001US-276994P.
PR 20-MAR-2001; 2001US-277239P.
PR 20-MAR-2001; 2001US-277321P.
PR 20-MAR-2001; 2001US-277327P.
PR 20-MAR-2001; 2001US-277338P.
PR 21-MAR-2001; 2001US-277791P.
PR 22-MAR-2001; 2001US-277833P.
PR 23-MAR-2001; 2001US-278152P.
PR 26-MAR-2001; 2001US-278894P.
PR 27-MAR-2001; 2001US-278999P.
PR 27-MAR-2001; 2001US-279036P.
PR 28-MAR-2001; 2001US-279344P.
PR 30-MAR-2001; 2001US-279995P.
PR 30-MAR-2001; 2001US-280233P.
PR 02-APR-2001; 2001US-280802P.
PR 02-APR-2001; 2001US-280822P.
PR 02-APR-2001; 2001US-280900P.
PR 04-APR-2001; 2001US-281194P.
PR 13-APR-2001; 2001US-283675P.
PR 30-APR-2001; 2001US-287424P.
PR 02-MAY-2001; 2001US-288066P.
PR 03-MAY-2001; 2001US-288342P.
PR 03-MAY-2001; 2001US-288528P.
PR 15-MAY-2001; 2001US-291190P.
PR 16-MAY-2001; 2001US-291099P.
PR 16-MAY-2001; 2001US-291240P.
PR 30-MAY-2001; 2001US-294485P.
PR 31-MAY-2001; 2001US-294889P.
PR 31-MAY-2001; 2001US-294899P.
PR 18-JUN-2001; 2001US-299027P.
PR 19-JUN-2001; 2001US-299303P.
PR 19-JUN-2001; 2001US-299310P.
PR 10-JUL-2001; 2001US-304354P.
PR 31-JUL-2001; 2001US-309198P.
PR 16-AUG-2001; 2001US-312903P.
PR 10-SEP-2001; 2001US-318462P.
PR 12-SEP-2001; 2001US-318770P.
PR 27-SEP-2001; 2001US-325430P.
PR 27-SEP-2001; 2001US-325681P.
PR 18-OCT-2001; 2001US-330380P.
PR 31-OCT-2001; 2001US-335301P.
PR 14-NOV-2001; 2001US-332172P.
PR 14-NOV-2001; 2001US-332271P.
PR 14-NOV-2001; 2001US-332272P.
PR 14-NOV-2001; 2001US-333184P.
PR 14-NOV-2001; 2001US-333272P.
PR 21-NOV-2001; 2001US-333272P.
PR 21-NOV-2001; 2001US-332094P.
PR 03-DEC-2001; 2001US-337426P.
PR 03-DEC-2001; 2001US-338092P.
PR 04-DEC-2001; 2001US-337185P.
PR 03-JAN-2002; 2002US-345705P.
PR 08-MAR-2002; 2002US-0093463.
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Caeman SJ;
PI Boldog FL, Li L, Zernusen BD, Tchernev VT, Gangolli EA;
PI Vernet CAM, Pena CE, Burgess CE, Liu X, Spytek KA, Gorman L;
PI Spaderna SK, Voss BZ, Malyankar UM, Anderson DW, Patutirajan M;
PI Miller CE, Taupier RJ, Padigaru M, Shenoy SG, Kekuda R, Gusev VY;
PI Pochart PF, Zhong M;
XX
DR WPI, 2002-732624/79.

```


PT diagnosing or treating NOVX-associated disorders e.g. diabetes, cancer,
PT Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic
PT disorders, and asthma
XX
PS Claim 16: Page 263-264; 619pp; English.
XX
XX The present invention relates to new isolated proteins (NOVX) and their
CC coding sequences (ABV9937-ABV9959 and ABP70049-ABP70149), where X is
CC any number from 1 to 48. The NOVX proteins and coding sequences are
CC useful in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, preferably a NOVX-associated disorder.
CC The NOVX coding sequences and proteins are useful for treating,
CC preventing or diagnosing diseases such as metabolic disorders, diabetes,
CC obesity, infectious disease, anorexia, cancer-associated cachexia,
CC cancer, neurodegenerative diseases, Alzheimer's disease, Parkinson's
CC disease, immune disorders, hematopoietic disorders, cardiovascular
CC disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic
CC disturbances associated with obesity, metabolic syndrome X or wasting
CC disorders associated with chronic diseases or various cancers. The NOVX
CC coding sequences and proteins may also be used as targets for the
CC identification of small molecules that modulate or inhibit e.g.
CC neurogenesis, cell differentiation, cell proliferation, hematopoiesis,
CC wound healing and angiogenesis, in gene therapy, in generation of
CC antibodies that bind immunospecifically to NOVX substances for use in
CC therapeutic or diagnostic methods.
XX
SQ Sequence 2563 BP; 371 A; 921 C; 857 G; 414 T; 0 other;
XX
Query Match 30.8%; Score 16; DB 25; Length 2563;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 33 CGCTCCGGCGCTCGCC 48
|||
Db 1679 CGCTCCGGCGCTCGCC 1694
|||
RESULT 16
AAK99609 standard; DNA; 2569 BP.
XX
XX AAK99609;
XX
DT 10-APR-2003 (first entry)
DE
XX MDDT related human DNA SEQ ID No 73.
XX
XX Cystostatic; antiatherosclerotic; osteopathic; antiarteriosclerotic;
KM hepatocytic; antiparasitic; antiallergic; antianaemic; antiasthmatic;
KM antithyroid; antinflammatory; antihelminthic; antidiabetic; nephrotropic;
KM ophthalmological; immunosuppressive; dermatological; antitumor;
KM antirheumatic; antiarthritic; antibacterial; virucide; fungicide;
KM antiparasitic; protozoacide; tranquiliser; vulnerary; anti-HIV;
KM neurotropic; neuroprotective; anticonvulsant; cerebroprotective;
KM neuroleptic; molecules for disease detection and treatment; MDDT;
KM immunogen; cancer; actinic keratosis; arteriosclerosis; atherosclerosis;
KM bursitis; cirrhosis; hepatitis; psoriasis; AIDS; rheumatoid arthritis;
KM adult respiratory distress syndrome; Addison's disease; allergy; anaemia;
KM asthma; osteoporosis; autoimmune; haemolytic anaemia; scleroderma;
KM autoimmune thyroiditis; Crohn's disease; atopic dermatitis;
KM diabetes mellitus; Graves' disease; glomerulonephritis;
KM systemic lupus erythematosus; systemic sclerosis; ulcerative colitis;
KM haemodialysis; uveitis; trauma; Alzheimer's; Pick disease;
KM Parkinson disease; amyotrophic lateral sclerosis; epilepsy; stroke;
KM Huntington's disease; multiple sclerosis; dementia;
KM extrapyramidal disorder; motor neuron disorder; central nervous system;
KM neuromuscular disorder; metabolic; endocrine; toxic myopathy;
KM periodic paralysis; mental disorder; human; gene; ds.
XX
XX Homo sapiens.
XX
OS
XX
PN MO200296951-A1.
XX

PD 05-DEC-2002.
XX
XX 24-MAY-2002; 2002WO-US16676.
PF
XX
XX 25-MAY-2001; 2001US-293723P.
PR 01-JUN-2001; 2001US-295257P.
PR 08-JUN-2001; 2001US-297220P.
PR 21-JUN-2001; 2001US-300526P.
PR 29-JUN-2001; 2001US-301874P.
PR 22-FEB-2002; 2002US-359413P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
PI Tang TY, Yue H, Baughn MR, Duggan BM, Warren BA, Bandman O,
PI Richardson TW, Barford N, Sanjamaala B, Becha SD, Yao MG, Yang J;
PI Tian UK, Hatalia AJA, Griffin JA, Swarnakar A, Elliott VS;
PI Recipon SA, Khan FA, Lee EA, Yue H, Lu DM, Walla NK;
PI Thangavelu K, Arvizu CS, Xu Y, Ison CH, Huang J, Ding L;
PI Honchell CD, Borowsky ML, Emerling BM, Peterson DP, Lu Y;
PI Ramkumar J, Mason PM, Zebairjadian Y, Azimzai Y, Stuve LL;
PI Kamigaki LL, Barroso I, Lee S, Kable AE;
XX
XX WPI: 2003-140448/13.
DR P-PSDB; AAO26256.
XX
PT Novel molecules for disease detection and treatment and polynucleotide
PT encoding them useful for diagnosing, preventing or treating cell
PT proliferative, autoimmune/inflammatory, neurological and developmental
PT disorders
XX
XX Claim 128; Page 255; 260pp; English.
XX
XX The invention relates to an isolated polypeptide chosen from molecules
CC for disease detection and treatment (MDDT), comprising a one of 39 114-
CC 1250 residue amino acid sequences, given in the specification, or a
CC biologically active or immunogenic fragment of the isolated polypeptide.
CC The isolated polypeptide is useful for screening a compound for
CC effectiveness as an agonist or antagonist of the isolated polypeptide.
CC The isolated polypeptide is also useful as an immunogen for preparing
CC polyclonal or monoclonal antibodies by hybridoma technology. The
CC isolated polypeptide and its encoding polynucleotide are useful for
CC diagnosis, treatment and prevention of cancer, actinic keratosis,
CC arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis,
CC psoriasis, AIDS, adult respiratory distress syndrome, Addison's disease,
CC allergies, anaemia, asthma, atherosclerosis, osteoporosis, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, Crohn's disease, atopic
CC dermatitis, diabetes mellitus, Graves' disease, glomerulonephritis,
CC rheumatoid arthritis, scleroderma, systemic lupus erythematosus,
CC bacterial, fungal, parasitic, protozoal, helminthic infections, trauma,
CC Alzheimer's and Pick disease, Parkinson disease, amyotrophic lateral
CC sclerosis, epilepsy, stroke, Huntington's disease, multiple sclerosis,
CC dementia, and other extrapyramidal disorder, motor neuron disorder, and
CC other developmental disorders of the central nervous system,
CC neuromuscular disorders, metabolic, endocrine and toxic myopathies,
CC periodic paralysis, mental disorders including mood, anxiety and
CC schizophrenic disorders, anaemia, renal tubular acidosis, epilepsy,
CC hypothyroidism, glaucoma, sensorineural hearing loss and cataract. This
CC polynucleotide sequence represents the DNA encoding a human MDDT protein
XX relating to the invention.
XX
SQ Sequence 2569 BP; 361 A; 928 C; 818 G; 461 T; 1 other;
XX
Query Match 30.8%; Score 16; DB 25; Length 2569;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 33 CGCTCCGGCGCTCGCC 48
|||
Db 1862 CGCTCCGGCGCTCGCC 1877
|||
RESULT 17

ABV99420	ID	ABV99420	standard; DNA; 3201 BP.
AC	XX	ABV99420;	
XX	XX	27-JAN-2003	(first entry)
XX	DE	Human NOV44a	coding sequence.
XX	KW	Human; anti-HIV; cytostatic; antidiabetic; antiasthmatic; cachexia; AIDS	
KW	KW	antiflammatory; cardiatic; haemostatic; neuroprotective; anorectic;	
KW	KW	neurotropic; immunosuppressive; osteopathic; antiparkinsonian; cancer;	
KW	KW	antifertility; cerebroprotective; gene therapy; NOV; NOV; fertility;	
KW	KW	metabolic disorder; diabetes; obesity; infectious disease; anorexia;	
KW	KW	neurodegenerative disease; Alzheimer's disease; Parkinson's disease;	
KW	KW	immune disorder; hematopoietic disorder; cardiovascular disorder;	
KW	KW	bronchial asthma; dyslipidemia; metabolic disturbance; neurogenesis;	
KW	KW	metabolic syndrome X; wasting disorder; cell differentiation;	
KW	KW	Single nucleotide polymorphism; SNP; cell proliferation; haematopoesis;	
KW	KW	wound healing; angiogenesis; gene; ds.	
XX	OS	Homo sapiens.	
XX	XX	Key	Location/Qualifiers
XX	XX	variation	replace(604,G)
XX	XX		/*tag= a
XX	XX		/standard_name= "Single nucleotide polymorphism"
XX	XX		replace(684,G)
XX	XX		/*tag= b
XX	XX	variation	/standard_name= "Single nucleotide polymorphism"
XX	XX		replace(1123,A)
XX	XX		/*tag= c
XX	XX		/standard_name= "Single nucleotide polymorphism"
XX	XX	variation	replace(1489,C)
XX	XX		/*tag= d
XX	XX		/standard_name= "Single nucleotide polymorphism"
XX	XX	MO200272771-A2.	
XX	XX	19-SEP-2002.	
XX	XX	08-MAR-2002; 2002MO-US07288.	
XX	XX	08-MAR-2001; 2001US-274101P.	
XX	XX	08-MAR-2001; 2001US-274194P.	
XX	XX	08-MAR-2001; 2001US-274281P.	
XX	XX	08-MAR-2001; 2001US-274322P.	
XX	XX	09-MAR-2001; 2001US-274849P.	
XX	XX	12-MAR-2001; 2001US-275235P.	
XX	XX	13-MAR-2001; 2001US-275578P.	
XX	XX	13-MAR-2001; 2001US-275579P.	
XX	XX	13-MAR-2001; 2001US-275601P.	
XX	XX	14-MAR-2001; 2001US-276000P.	
XX	XX	16-MAR-2001; 2001US-276776P.	
XX	XX	19-MAR-2001; 2001US-276994P.	
XX	XX	20-MAR-2001; 2001US-277239P.	
XX	XX	20-MAR-2001; 2001US-277321P.	
XX	XX	20-MAR-2001; 2001US-277327P.	
XX	XX	20-MAR-2001; 2001US-277338P.	
XX	XX	21-MAR-2001; 2001US-277791P.	
XX	XX	22-MAR-2001; 2001US-277833P.	
XX	XX	23-MAR-2001; 2001US-278152P.	
XX	XX	26-MAR-2001; 2001US-278894P.	
XX	XX	27-MAR-2001; 2001US-278999P.	
XX	XX	28-MAR-2001; 2001US-279034P.	
XX	XX	30-MAR-2001; 2001US-279995P.	
XX	XX	30-MAR-2001; 2001US-280233P.	
XX	XX	02-APR-2001; 2001US-280802P.	
XX	XX	02-APR-2001; 2001US-280822P.	
XX	XX	02-APR-2001; 2001US-280900P.	
XX	XX	04-APR-2001; 2001US-281194P.	
XX	XX	13-APR-2001; 2001US-281675P.	

PR	30-APR-2001	2001US-287424P
PR	02-MAY-2001	2001US-288066P
PR	03-MAY-2001	2001US-288342P
PR	03-MAY-2001	2001US-288528P
PR	15-MAY-2001	2001US-291190P
PR	16-MAY-2001	2001US-291099P
PR	16-MAY-2001	2001US-291240P
PR	30-MAY-2001	2001US-294485P
PR	31-MAY-2001	2001US-294889P
PR	31-MAY-2001	2001US-294899P
PR	18-JUN-2001	2001US-299027P
PR	19-JUN-2001	2001US-299303P
PR	19-JUN-2001	2001US-299310P
PR	10-JUL-2001	2001US-304354P
PR	31-JUL-2001	2001US-309198P
PR	16-AUG-2001	2001US-312903P
PR	10-SEP-2001	2001US-318462P
PR	12-SEP-2001	2001US-325430P
PR	27-SEP-2001	2001US-325681P
PR	18-OCT-2001	2001US-330380P
PR	31-OCT-2001	2001US-335301P
PR	14-NOV-2001	2001US-332172P
PR	14-NOV-2001	2001US-332271P
PR	14-NOV-2001	2001US-332272P
PR	14-NOV-2001	2001US-333184P
PR	14-NOV-2001	2001US-333272P
PR	21-NOV-2001	2001US-333094P
PR	03-DEC-2001	2001US-337426P
PR	03-DEC-2001	2001US-338092P
PR	04-DEC-2001	2001US-337185P
PR	03-JAN-2002	2002US-345705P
PR	08-MAR-2002	2002US-0093463
PA	(CURA-) CURAGEN CORP.	
XX		
PI	Rastelli L, Mezes PD, Smithson G, Guo X, Gerlach V, Casman SJ;	
PI	Bojdov Fu, Li L, Zernusen BD, Tchernev VT, Gangolli EA;	
PI	Vernet CM, Pena CE, Burgess CE, Liu X, Spyrek KA, Gorman L;	
PI	Spaderna SK, Voss EZ, Maljenkar UM, Anderson DW, Ratturajan M;	
PI	Miller CE, Taupier RJ, Padigar M, Shenoy SG, Kekuda R, Guise VY;	
PI	Pochart P, Zhong M;	
XX		
DR	WPI: 2002-732824/79.	
DR	P-PSDB; ABP70142.	
XX		
PT	New NOXV polypeptides and polynucleotides, useful for preventing,	
PT	diagnosing or treating NOXV-associated disorders e.g. diabetes, cancer,	
PT	Alzheimer's disease, dyslipidemias, obesity, immune or hematopoietic	
PT	disorders, and asthma -	
XX		
PS	Claim 16; Page 261-262; 61pp: English.	
XX		
CC	The present invention relates to new isolated proteins (NOXV) and their	
CC	coding sequences (ABV99327-ABV99595 and ABP70049-ABP70149), where X is	
CC	any number from 1 to 48. The NOXV proteins and coding sequences are	
CC	useful in the manufacture of a medicament for treating a syndrome	
CC	associated with a human disease, preferably a NOXV-associated disorder.	
CC	The NOXV coding sequences and proteins are useful for treating	
CC	preventing or diagnosing diseases such as metabolic disorders, diabetes	
CC	cancer, infectious disease, anorexia, cancer-associated cachexia,	
CC	obesity, neurodegenerative diseases, Alzheimer's disease, Parkinson's	
CC	disease, immune disorders, haematopoietic disorders, cardiovascular	
CC	disorders, fertility, bronchial asthma, AIDS, dyslipidemia, metabolic	
CC	disturbances associated with obesity, metabolic syndrome X or wasting	
CC	disorders associated with chronic diseases or various cancers. The NOXV	
CC	coding sequences and proteins may also be used as targets for the	
CC	identification of small molecules that modulate or inhibit e.g.	
CC	neurogenesis, cell differentiation, cell proliferation, haematopoiesis,	
CC	wound healing and angiogenesis, in gene therapy, in generation of	
CC	antibodies that bind immunospecifically to NOXV substances for use in	
XX	therapeutic or diagnostic methods.	

Sequence 3201 BP; 534 A; 1174 C; 986 G; 507 T; 0 other;

Query Match 30.8%; Score 16; DB 24; Length 3201;

Best Local Similarity 100.0%; Pred. NO. 26;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 33 CGCTCCGCGCTCGCC 48

Db 1966 CGCTCCGCGCTCGCC 1981

RESULT 18

AAFI6203/C

ID AAFI6203 standard; cDNA; 3947 BP.

XX AAFI6203;

DT 13-MAR-2001 (first entry)

DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:638.

XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
neuroprotective; cytosolic; immunomodulatory; muscular;
vulnary; gastrointestinal; nephrotoxic; antineoplastic; gynaecological;
antibacterial; gene therapy; neural; immune; reproductive; renal;
gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
wound; infectious disease; ss.

OS Homo sapiens.

PN WO200055174-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US05988.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

PI Rosen CA, Ruben SM;

DR WPI; 2000-587513/55.

DR P-PSDB; AAB57000.

PT Prostate cancer associated gene sequences, referred to as prostate

PT cancer antigens, useful for treatment, prevention, and diagnosis of

PT disorders such as prostate cancer -

XX Claim 1; Page 1079-1080; 2338pp; English.

XX AAFI556 to AAFI505 encode the human prostate cancer associated

CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.

CC The prostate cancer antigens can have neuroprotective, cytosolic,

CC cardioprotective, immunomodulatory, muscular, vulnary, gastrointestinal,

CC nephrotoxic, antineoplastic, gynaecological and antibacterial activities,

CC and can be used in gene therapy. The prostate cancer antigen

CC polynucleotides may be used for detection of prostate cancer, chromosome

CC identification, as chromosome markers, and for numerous other diagnostic

CC or research purposes. The prostate cancer antigens may be used to treat

CC disorders such as neural, immune, muscular, reproductive,

CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

CC disorders, wounds, and infectious diseases. AAFI506 to AAFI514 to

CC AAB57303 represent sequences used in the exemplification of the present

CC invention.

XX Sequence 3947 BP; 1067 A; 979 C; 983 G; 910 T; 8 other;

XX Query Match 30.8%; Score 16; DB 21; Length 3947;

XX Best Local Similarity 100.0%; Pred. NO. 26;

XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 CTCATCGCTCTGTCG 34

Db 72 CTCATCGCTCTGTCG 57

RESULT 19

AAH34839/C

ID AAH34839 standard; cDNA; 3953 BP.

XX AAH34839;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen encoding cDNA SEQ ID NO:1921.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma; ss.

OS Homo sapiens.

PN WO200122920-A2.

PD 05-APR-2001.

PF 28-SEP-2000; 2000MO-US26524.

PR 29-SEP-1999; 99US-0157137.

PR 03-NOV-1999; 99US-0163280.

PA (HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, Barash SC, Birse CE, Rosen CA;

PI WPI; 2001-235357/24.

DR P-PSDB; AAG75434.

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

PT useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 1; Page 3431-3432; 9803pp; English.

XX AAH2943 to AAH37195 and AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where

CC the proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytosolic activity and can be used in gene

CC therapy and vaccine production. N and P may be used in the prevention,

CC diagnosis and treatment of diseases associated with inappropriate P

CC expression. For example, N and P may be used to treat disorders

CC associated with decreased expression by rectifying mutations or deletions

CC in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patient's own production of P.

CC Additionally, N may be used to produce the colon cancer-associated P,

CC by inserting the nucleic acids into a host cell and culturing the cell

CC to express the proteins. N and P can be used in the prevention, diagnosis

CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204

CC and AAB7789 represent sequences used in the exemplification of the

CC present invention.

CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for

CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 3953 BP; 1055 A; 972 C; 986 G; 930 T; 10 other;

XX Query Match 30.8%; Score 16; DB 22; Length 3953;

XX Best Local Similarity 100.0%; Pred. NO. 26;

XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX Qy 19 CTCATCGCTCTGTCG 34

XX Db 72 CTCATCGCTCTGTCG 57

XX RESULT 20


```

ABLI5493
ID ABLI5493 standard; cDNA; 6507 BP.
XX
XX
AC ABLI5493;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 40961.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB71390.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 40961; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX sequences (ABLI01840-ABLI6175) and the encoded proteins
XX (AAB57737-ABR72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6507 BP; 1745 A; 1888 C; 1712 G; 1162 T; 0 other;
SQ
Query Match 30.8%; Score 16; DB 23; Length 6507;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 21 CATCGCTCTGCTGCT 36
Db 2990 CATCGCTCTGCTGCT 3005

```

```

RESULT 21
AAS84163/c
ID AAS84163 standard; cDNA; 7434 BP.
XX
XX
AC AAS84163;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #19967.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS

```

```

XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG19976.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID NO 19967; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 7434 BP; 1795 A; 2099 C; 1892 G; 1646 T; 2 other;
SQ
Query Match 30.8%; Score 16; DB 23; Length 7434;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 19 CTCATCGCTCTGCTG 34
Db 2228 CTCATCGCTCTGCTG 2213

```

```

RESULT 22
ABLI4186/c
ID ABLI4186 standard; cDNA; 15435 BP.
XX
XX
AC ABLI4186;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 37040.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN

```

XX 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PF 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX P-PSDB; ABB70083.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 37040; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU1840-ABU16175) and the encoded proteins
CC (ABBS7737-ABBS7072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 15435 BP; 4735 A; 3331 C; 2995 G; 4374 T; 0 other;
Query Match 30.8%; Score 16; DB 23; Length 15435;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 15 CGCTCTCATCGCTCTC 30
Db 2140 CGCTCTCATCGCTCTC 2125
RESULT 23
AAD52898
ID AAD52898 standard; DNA; 47999 BP.
XX
AC AAD52898;
XX
DT 14-MAY-2003 (first entry)
XX
DE Human tweety homologue 2 (TTYH2) gene.
XX
KM Human; tweety homologue 2; TTYH2; therapy; cancer; tumour; cytostatic;
KM diagnostic marker; gene; ds.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT CDS 1946..45753
FT /tag= a
FT /product= "Human TTYH2 protein"
FT exon 1936..2074
FT /tag= b
FT /number= 1
FT intron 2075..10376
FT /tag= c
FT exon 10377..10549
FT /tag= d
FT /number= 2
FT intron 10550..16622
FT /tag= e

FT exon 16623..16734
FT /tag= f
FT /number= 3
FT intron 16735..22323
FT /tag= g
FT exon 23224..23444
FT /tag= h
FT /number= 4
FT intron 23445..28299
FT /tag= i
FT exon 28300..28395
FT /tag= j
FT /number= 5
FT intron 28396..28902
FT /tag= k
FT exon 28903..28975
FT /tag= l
FT /number= 6
FT intron 28976..35372
FT /tag= m
FT exon 35373..35442
FT /tag= n
FT /number= 7
FT intron 35443..35705
FT /tag= o
FT exon 35706..35761
FT /tag= p
FT /number= 8
FT intron 35762..36266
FT /tag= q
FT exon 36267..36359
FT /tag= r
FT /number= 9
FT intron 36360..36591
FT /tag= s
FT exon 36592..36684
FT /tag= t
FT /number= 10
FT intron 36685..368529
FT /tag= u
FT exon 368530..36872
FT /tag= v
FT /number= 11
FT intron 36873..39376
FT /tag= w
FT exon 39377..39562
FT /tag= x
FT /number= 12
FT intron 39563..40050
FT /tag= y
FT exon 40051..40129
FT /tag= z
FT /number= 13
FT intron 40130..45672
FT /tag= aa
FT exon 45673..47158
FT /tag= ab
FT /number= 14
XX
XX WO200292629-A1.
XX
XX 21-NOV-2002.
XX
XX 14-MAY-2002; 2002WO-AU00591.
XX
XX 14-MAY-2001; 2001AU-0004971.
XX
XX (UYOU-) UNIV QUEENSLAND TECHNOLOGY.
XX
XX Clements JA;
XX
XX WPI; 2003-129264/12.
XX
XX P-PSDB; AAE34613.
DR

XX New human tweety homolog 2 polypeptides and polynucleotides, useful for
PT producing an antigen-binding molecule that is immuno-interactive with
PT the polypeptide or as diagnostic markers for cancers -
XX
XX
PS Claim 10; Page 128-156; 176pp; English.
XX
XX The invention relates to human tweety homologue 2 (TYH2) polypeptide and
CC polynucleotide sequence. TYH2 is useful for producing an antigen-binding
CC molecule that is immuno-interactive with the polypeptide. The agent is
CC useful for manufacturing a medicament for restoring a normal level and/or
CC functional activity of TYH2 expression in a patient, and for treating or
CC preventing cancer or tumour. TYH2 sequences may also be used to provide
CC both drug targets and regulators to promote or inhibit one or more
CC activities, and to provide diagnostic markers for cancers. The present
CC sequence is human TYH2 gene.
XX
SO Sequence 47999 BP; 11178 A; 12727 C; 12874 G; 11209 T; 11 other;
Query Match 30.8%; Score 16; DB 25; Length 47999;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 CGCTCTCATCGCTCTC 30
DB 1247 CGCTCTCATCGCTCTC 1262
RESULT 24
ABL15492
ID ABL15492 standard; cDNA; 59967 BP.
XX
XX ABL15492;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 40958.
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 40958.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
XX P-PSDB; ABB11389.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT gene from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 40958; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL1840-ABL16175) and the encoded proteins
CC (ABB5737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SO Sequence 59967 BP; 17305 A; 12463 C; 12291 G; 17908 T; 0 other;
Query Match 30.8%; Score 16; DB 23; Length 59967;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 CATCGCTCTCGTCTC 36
DB 4640 CATCGCTCTCGTCTC 4655
RESULT 25
ABQ81844/C
ID ABQ81844 standard; DNA; 349980 BP.
XX
XX ABQ81844;
XX
XX 19-NOV-2002 (first entry)
XX
XX Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1100.
DE Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW anti-diarrhetic; antibacterial; inhibitor of Salmonella; detection;
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
XX Bifidobacterium longum.
OS
XX Synthetic.
XX
XX EP1227152-A1.
XX
XX 31-JUL-2002.
XX
XX 30-JAN-2001; 2001EP-0102050.
XX
XX 30-JAN-2001; 2001EP-0102050.
PR
XX (NEST) SOC PROD NESTLE SA.
XX
XX WPI; 2002-668397/72.
XX
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful
PT as a probe or primer for detecting and/or identifying Bifidobacterium
PT longum in a biological sample -
XX
XX Disclosure; SEQ ID 1100; 80pp; English.
XX
XX The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC least 90% identity or which hybridises with the sequences given in
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding
CC a fusion protein, comprising a sequence selected from 1097 sequences
CC given in ABP65258 to ABP6354 ligated in frame to a polynucleotide
CC encoding a heterologous polypeptide. (I) has anti-diarrhetic and
CC antibacterial activities, and can be used as an inhibitor of Salmonella.
CC (I) (which is a probe) is useful for the detection and/or identification
CC of Bifidobacterium longum in a biological sample. A carrier containing
CC the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618)
CC can be used for preventing and/or treating diarrhoea brought about by
CC pathogenic bacteria and/or rotavirus. The carrier is a food composition
CC selected from milk, yogurt, curd, cheese, fermented milks, milk based
CC fermented products, ice-creams, fermented cereal based products, milk
CC based powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
CC (I) is useful in DNA arrays or chips to carry out analysis of the
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent
CC Bifidobacterium related nucleotide sequences given in the Sequence

CC Listing from the present invention but not mentioned further within the
CC specification.
CC N.B. The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 349980 BP; 69975 A; 105045 C; 104394 G; 70566 T; 0 other;
Query Match 30.8%; Score 16; DB 24; Length 349980;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 TCATCGCTCTCGTCGC 35
Db 253393 TCATCGCTCTCGTCGC 253378
RESULT 26
ABL74833
ID ABL74833 standard; cDNA; 214 BP.
XX
AC ABL74833;
XX
DT 14-MAY-2002 (first entry)
XX
DE Corn tassel-derived polynucleotide (cdps) SEQ ID NO:4207.
XX
XX Corn: corn tassel-derived polynucleotide; cdps: hybrid breeding; CDPS:
XX inheritance; characteristic; growth; development; disease resistance;
XX environmental adaptability; quality; yield; molecular marker;
XX multi-gene trait; plant breeding; corn tassel; gene; ss.
XX
XX Zea mays.
XX
XX US2001051335-A1.
XX
XX 13-DEC-2001.
XX
XX 16-APR-1999; 99US-0294093.
XX
XX 21-APR-1998; 98US-082567P.
XX
XX (LALGU/) LALGUDI R V.
XX (ITOL/) ITO L Y.
XX (SHER/) SHERMAN B K.
XX
XX Lalgudi RV, Ito LY, Sherman BK;
XX
XX WPI: 2002-163647/21.
XX
XX Novel purified corn tassel-derived polynucleotide useful for
XX determining altered gene expression, to recover regulatory elements and
XX to follow inheritance of desirable characteristics through hybrid
XX breeding programs -
XX
XX Claim 1; SEQ ID 4207; 201bp; English.
XX
XX The present sequence describes a purified corn tassel-derived
XX polynucleotide sequence (cdps) comprising a nucleic acid sequence
XX selected from those given in ABL70627 to ABL76833. The cdps sequences
XX can be used for determining altered gene expression, to recover
XX regulatory elements and to follow inheritance of desirable
XX characteristics through hybrid breeding programs. (I) are also useful
XX in the evaluation, and alteration of desired characteristics associated
XX with growth and development, disease resistance, environmental
XX adaptability, quality and yield, and as molecular markers for studying
XX inheritance of multi-gene traits in a plant breeding program. (I) can be
XX used to produce a tassel-specific profile of gene transcription, a
XX transcript image, to clone regulatory elements for use in transformation
XX vectors, to express a polypeptide, to identify, isolate or extend
XX identical or related corn tassel nucleic acid sequences from DNA
XX libraries, in nucleic acid hybridisation or amplification technologies,

CC as query sequences to determine homology of known sequences, as probe
CC for use in Southern or Northern hybridisation, and to identify the
CC presence of and/or to determine the degree of similarity between two
CC (or more) nucleic acid sequences.
XX
SQ Sequence 214 BP; 68 A; 31 C; 50 G; 62 T; 3 other;
Query Match 28.8%; Score 15; DB 24; Length 214;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGTCTCTCAC 15
Db 182 ATGCAGTCTCTCAC 196
RESULT 27
ABX33704/C
ID ABX33704 standard; cDNA; 264 BP.
XX
AC ABX33704;
XX
DT 11-FEB-2003 (first entry)
XX
XX Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #15761.
XX
XX Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation;
XX cellular fucosylation; glycoconjugate fucosylation; transplant rejection;
XX arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
XX complex carbohydrate; gene replacement therapy; immunosuppressive;
XX anti-inflammatory; antiarthritic; antibacterial; cerebroprotective;
XX antiasthmatic; vasotropic.
XX
XX Homo sapiens.
XX
XX US2002110548-A1.
XX
XX 15-AUG-2002.
XX
XX 11-JUN-2001; 2001US-0878574.
XX
XX 22-NOV-1996; 96US-0753233.
XX 03-DEC-1997; 97US-0984246.
XX 09-SEP-1998; 98US-0149674.
XX 14-JUN-1999; 99US-0333177.
XX
XX (GENW) GENETICS INST INC.
XX
XX Sullivan F, Kriz R, Kumar R;
XX
XX WPI: 2003-066673/06.
XX
XX New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)
XX peptide, for manufacturing complex carbohydrates, or as targets for
XX screening GM4,6D antagonists for treating e.g. arthritis, or transplant
XX rejection -
XX
XX Disclosure; SEQ ID NO 15763; 6pp; English.
XX
XX The invention relates to a composition comprising a human GDP-mannose
XX 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying
XX GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation
XX in a mammalian subject and for treating or ameliorating diseases affected
XX by the level of cellular fucosylation or diseases affected by the
XX fucosylation of glycoconjugates. These diseases include arthritis,
XX transplant rejection, asthma, sepsis, reperfusion injury, stroke or
XX infection. The GM4,6D peptide or a polynucleotide encoding it is also
XX useful for manufacturing complex carbohydrates and as targets for
XX screening small molecule antagonists of the activity of the enzyme. The
XX polynucleotide is useful in developing an assay for defects in the
XX enzyme, as well as in gene replacement therapy. Sequences
XX ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding
XX human GM4,6D peptides of the invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at segdata.uspto.gov/sequence.html.

XX Sequence 264 BP; 66 A; 64 C; 71 G; 63 T; 0 other;

QY Query Match 28.8%; Score 15; DB 25; Length 264;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 26 CTCTCGTCGCTGCCG 40
111 CTCTCGTCGCTGCCG 97

RESULT 28

AB217641
ID AB217641 standard; cDNA; 335 BP.

AC AB217641;

DT 23-JAN-2003 (first entry)

DE S2 subtraction library cancer related clone SEQ ID NO:67.

KW Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
KW immune response; virology; immunology; microbiology; molecular biology;
KW recombinant DNA technology; gene; ss.

XX Homo sapiens.

PN WO200278516-A2.

PD 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US10421.

PR 30-MAR-2001; 2001US-280255P.

PR 28-AUG-2001; 2001US-315563P.

PR 09-JAN-2002; 2002US-347313P.

PA (CORI-) CORIXA CORP.

PI Wang T, Wang S, Bangur CS, Gaiger A;

XX WPI; 2003-058387/05.

PT New immunogenic polynucleotides or polypeptides useful for diagnosing,
PT preventing and treating cancer expressing CT or CP mRNA antigens, and
PT in virology, immunology, microbiology, molecular biology and
PT recombinant DNA techniques -

PS Claim 1; SEQ ID 67; 207bp; English.

CC AB017575 to AB020506 represent isolated polynucleotide (I) sequences, and
CC ABP54446 to ABP54472 represent protein (II) sequences, from the present
CC invention. (I) and (II) have cytoskeletal activity and can be used in gene
CC therapy and vaccines. (I), (II), antibodies and compositions from the
CC present invention are useful for diagnosing, preventing and treating
CC cancer, which expresses CT or CP mRNA antigens. They are useful for
CC stimulating immune response. They can also be useful in virology,
CC immunology, microbiology, molecular biology and recombinant DNA
CC techniques.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 335 BP; 77 A; 96 C; 96 G; 66 T; 0 other;

QY Query Match 28.8%; Score 15; DB 25; Length 335;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 CTCTCATCGCTCTCG 31
111 CTCTCATCGCTCTCG 151
Db 137 CTCTCATCGCTCTCG 151

RESULT 29

ABX19200/c
ID ABX19200 standard; cDNA; 336 BP.

AC ABX19200;

DT 10-FEB-2003 (first entry)

DE Human GDP-mannose 4,6-dehydratase (GM4, 6D) DNA #1257.

KW Human; GDP-mannose 4,6-dehydratase; GM4, 6D; gene; ss; inflammation;
KW cellular fucosylation; glycoconjugate fucosylation; transplant rejection;
KW arthritis; asthma; sepsis; reperfusion injury; stroke; infection;
KW complex carbohydrate; gene replacement therapy; immunosuppressive;
KW antiinflammatory; antibacterial; cerebroprotective;
KW antiasthmatic; vasotropic.

XX Homo sapiens.

PN US2002110548-A1.

PD 15-AUG-2002.

PF 11-JUN-2001; 2001US-0878574.

PR 22-NOV-1996; 96US-0753233.

PR 03-DEC-1997; 97US-0984246.

PR 09-SEP-1998; 98US-0149674.

PR 14-JUN-1999; 99US-0333177.

PA (GEMV) GENETICS INST INC.

PI Sullivan F, Kriz R, Kumar R;

XX WPI; 2003-066673/06.

PT New composition comprising GDP-mannose 4,6-dehydratase (GM4, 6D)
PT peptide, for manufacturing complex carbohydrates, or as targets for
PT screening GM4, 6D antagonists for treating e.g. arthritis, or transplant
PT rejection -

PS Disclosure; SEQ ID NO 1259; 6bp; English.

CC The invention relates to a composition comprising a human GDP-mannose
CC 4,6-dehydratase (GM4, 6D) peptide. The peptide is useful for identifying
CC GM4, 6D inhibitors. GM4, 6D inhibitors are useful for reducing inflammation
CC in a mammalian subject and for treating or ameliorating diseases affected
CC by the level of cellular fucosylation or diseases affected by the
CC fucosylation of glycoconjugates. These diseases include arthritis,
CC transplant rejection, asthma, sepsis, reperfusion injury, stroke or
CC infection. The GM4, 6D peptide or a polynucleotide encoding it is also
CC useful for manufacturing complex carbohydrates and as targets for
CC screening small molecule antagonists of the activity of the enzyme. The
CC polynucleotide is useful in developing an assay for defects in the
CC enzyme, as well as in gene replacement therapy. Sequences
CC ABX17942-ABX17944 and ABX17947-ABX3716 represent DNA molecules encoding
CC human GM4, 6D peptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at segdata.uspto.gov/sequence.html.

XX Sequence 336 BP; 115 A; 49 C; 80 G; 92 T; 0 other;

QY Query Match 28.8%; Score 15; DB 25; Length 336;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 67 CTCTCAGCTCTCAT 53

RESULT 30
ID AAS31315 standard; cDNA; 357 BP.
XX AAS31315;
AC
XX
XX 04-DEC-2001 (first entry)
XX
XX
DE Human cDNA encoding a novel extracellular matrix protein, Seq ID No 129.
XX
XX
KW Human; secreted extracellular matrix protein; ss; immunomodulatory;
KW Anti-HIV; antineoplastic; antirheumatic; antisclerotic; cardiac; vascular;
KW cerebroprotective; chondrolytic; antimicrobial; ophthalmic; cytoskeletal;
KW antialzheimers; immune/autoimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW cancers; hyperproliferative disorder; breast neoplasm; melanoma;
KW Sezary syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antisense; food additive.
XX
XX Homo sapiens.
XX
XX WO200155368-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01348.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225477.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236372.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249254.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465572/50.
DR P-PSDB; AAU19744.
XX

PT Nucleic acid molecules encoding human secreted extracellular matrix
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers -
XX
XX Claim 1; SEQ ID No 129; 577pp; English.

XX The invention relates to isolated nucleic acid molecules encoding
CC novel human secreted extracellular matrix proteins (SPs). The
CC polynucleotides and proteins are used to prevent, treat a medical
CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. For example, disorders associated with decreased
CC expression of SPs. The SP polynucleotide or a vector expressing them may
CC be administered to treat diseases by gene therapy. Antisense molecules
CC may be administered to down regulate expression of SPs by binding with
CC the cells own genes and preventing their expression. The polynucleotides
CC may also be used as DNA probes in diagnostic assays. The SPs may also be
CC used as antigens to produce antibodies and to identify modulators
CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and
CC antagonists may also be used to down regulate expression and activity of
CC SP and as diagnostic agents for detecting the presence of SPs in samples.
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac
CC arrest, tachycardia and angina), infections caused by bacteria, viruses
CC and fungi and ocular disorders (e.g. corneal infections). Other uses
CC include wound healing, maintenance of organs before transplantation,

Query Match 28.8%; Score 15; DB 22; Length 357;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGTCTCTCAG 15
|||||
Db 102 ATGCAGTCTCTCAG 116

RESULT 31
ABK43452
ID ABK43452 standard; cDNA; 357 BP.

XX
AC ABK43452;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding novel central nervous system protein #32.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX WO20015318-A2.
PN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX

PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230417.
PR 06-SEP-2000; 2000US-0230418.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.

```

PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.

```

```

PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251088.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX
XX DR WPI, 2001-581633/65.
XX DR P-PSDB; AAU87122.
XX
XX PT New isolated nucleic acid encoding a protein for diagnosing,
XX PT preventing, treating or ameliorating medical conditions and used as
XX PT food additives or preservatives -
XX
XX PS Claim 1; SEQ ID NO 42; 837bp; English.
XX
XX
XX CC The invention describes an isolated nucleic acid molecule (I) encoding a
XX CC novel central nervous system protein. (I) and polypeptides (II) encoded
XX CC by (I), are used to treat a medical conditions and in diagnosis of a
XX CC pathological condition. Disorders which are diagnosed or treated include
XX CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,
XX CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
XX CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
XX CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
XX CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
XX CC adenocarcinomas and irritable bowel syndrome, reproductive system
XX CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
XX CC leukaemia, disorders involving neovascularisation e.g. malignancies,
XX CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
XX CC acute kidney failure and blood related disorders e.g. myocardial
XX CC infarction. The polypeptides can also be used to aid wound healing and
XX CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
XX CC maintain organs before transplantation, for supporting cell culture of
XX CC primary tissues, to regenerate tissues and in chemotaxis. The
XX CC polypeptides can also be used as a food additive or preservative to
XX CC increase or decrease storage capabilities, fat content, lipid, protein,
XX
XX
XX Query Match 28.8%; Score 15; DB 23; Length 357;
XX Best Local Similarity 100.0%; Pred. No. 94;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 ATGAGTTCTCTCAC 15
XX |||||
XX DB 102 ATGAGTTCTCTCAC 116
XX
XX RESULT 32
XX ABO66639
XX ID ABO66639 standard; cDNA; 357 BP.
XX AC ABO66639;
XX XX
XX DT 23-AUG-2002 (first entry)
XX XX
XX DE Human polynucleotide SEQ ID NO 129.
XX XX

```


KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; anti-inflammation; anti-HIV; antibacterial; cancer;
 KW antiparkinsonian; antischistosomal; antianemic; antidiabetic; cancer;
 KW antihemorrhagic; hepatotropic; cerebroprotective; anti-inflammatory;
 KW antileukemic; antidiabetic; antileukemic; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotoxic; gene therapy; vaccine;
 KW gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN US2002042386-A1.
 XX
 PD 11-APR-2002.
 XX
 PF 17-JAN-2001; 2001US-0764870.
 XX
 PR 31-JAN-2000; 2000US-179065P.
 PR 04-FEB-2000; 2000US-160628P.
 PR 28-JUN-2000; 2000US-214886P.
 PR 07-JUL-2000; 2000US-216647P.
 PR 07-JUL-2000; 2000US-216880P.
 PR 11-JUL-2000; 2000US-217487P.
 PR 11-JUL-2000; 2000US-217496P.
 PR 14-JUL-2000; 2000US-218290P.
 PR 26-JUL-2000; 2000US-220963P.
 PR 26-JUL-2000; 2000US-220964P.
 PR 14-AUG-2000; 2000US-224518P.
 PR 14-AUG-2000; 2000US-224519P.
 PR 14-AUG-2000; 2000US-225267P.
 PR 14-AUG-2000; 2000US-225268P.
 PR 14-AUG-2000; 2000US-225270P.
 PR 14-AUG-2000; 2000US-225447P.
 PR 14-AUG-2000; 2000US-225757P.
 PR 14-AUG-2000; 2000US-225758P.
 PR 22-AUG-2000; 2000US-226868P.
 PR 30-AUG-2000; 2000US-228924P.
 PR 01-SEP-2000; 2000US-229287P.
 PR 01-SEP-2000; 2000US-229343P.
 PR 01-SEP-2000; 2000US-229344P.
 PR 01-SEP-2000; 2000US-229345P.
 PR 05-SEP-2000; 2000US-229509P.
 PR 05-SEP-2000; 2000US-229511P.
 PR 08-SEP-2000; 2000US-231413P.
 PR 21-SEP-2000; 2000US-234227P.
 PR 21-SEP-2000; 2000US-234274P.
 PR 25-SEP-2000; 2000US-234997P.
 PR 27-SEP-2000; 2000US-235834P.
 PR 29-SEP-2000; 2000US-236327P.
 PR 29-SEP-2000; 2000US-236367P.
 PR 29-SEP-2000; 2000US-236368P.
 PR 29-SEP-2000; 2000US-236369P.
 PR 29-SEP-2000; 2000US-236370P.
 PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237037P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237039P.
 PR 02-OCT-2000; 2000US-237040P.
 PR 13-OCT-2000; 2000US-239935P.
 PR 20-OCT-2000; 2000US-240960P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 01-NOV-2000; 2000US-244617P.
 PR 17-NOV-2000; 2000US-249299P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251868P.
 PR 08-DEC-2000; 2000US-251869P.
 XX
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PT Rosen CA, Ruben SM, Barash SC;

XX
 DR WPI; 2002-470713/50.
 DR P-PSDB; ABP47964.
 XX
 PT New nucleic acid encoding human proteins, useful for diagnosis.
 PT treatment and prevention of e.g. osteoporosis, also related
 PT polypeptides and antibodies
 XX
 PS Claim 1; SEQ ID NO 129; 235pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AB066521-AB066785) and proteins
 CC (ABP47846-ABP48110) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from USPTO at seqdata.uspto.gov/sequence.html?docID=99990764870.
 XX
 SQ Sequence 357 BP; 104 A; 68 C; 86 G; 93 T; 6 other;
 Query Match 28.8%; Score 15; DB 24; Length 357;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 ATGCAGTTCTCTCAC 15
 Db 102 ATGCAGTTCTCTCAC 116
 RESULT 33
 AAS76408/c
 ID AAS76408 standard; cDNA; 416 BP.
 XX
 AC AAS76408;
 XX
 DT 13-FEB-2002. (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #12212.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PT Dimanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG12221.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 1; SEQ ID No 12212; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

Sequence 416 BP; 74 A; 118 C; 128 G; 96 T; 0 other;

Query Match 28.8%; Score 15; DB 23; Length 416;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 CTCATCGCTCTCGTC 33
|||
Db 172 CTCATCGCTCTCGTC 158

RESULT 34

AA219396/C
ID AA219396 standard; cDNA; 423 BP.

XX
AC AA219396;

DT 05-NOV-1999 (first entry)

XX M. tuberculosis antigen 5' L5ER-5 cDNA sequence.

XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;

KW immunotherapy; diagnosis; immunisation; vaccine; infection;

KW immune response; skin test; ss.

XX Mycobacterium tuberculosis.

PN WO9942076-A2.

XX 26-AUG-1999.

PD 17-FEB-1999; 99WO-US03268.

PR 05-MAY-1998; 98US-0072967.

PR 18-FEB-1998; 98US-0025197.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;

XX WPI; 1999-527409/44.
DR P-PSDB; AAY39201.

XX New antigens from Mycobacterium tuberculosis useful in diagnostic
PT skin tests and protective or therapeutic vaccines or compositions

XX
PS Claim 11; Page 229; 29pp; English.

XX
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptide fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AA219249 to AA219460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX

Sequence 423 BP; 95 A; 128 C; 129 G; 71 T; 0 other;

Query Match 28.8%; Score 15; DB 20; Length 423;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 CGTCGCTGCCGCCCT 44
|||
Db 26 CGTCGCTGCCGCCCT 12

RESULT 35

AA219184/C
ID AA219184 standard; cDNA; 423 BP.

XX
AC AA219184;

DT 05-NOV-1999 (first entry)

XX M. tuberculosis recombinant antigen cDNA encoding 5' L5ER-5.

DE Antigen; diagnosis; detection; infection; antibody; immunisation;

XX vaccine; immunity; ss.

KW Mycobacterium tuberculosis.

XX WO9942118-A2.

PN 26-AUG-1999.

PD 17-FEB-1999; 99WO-US03265.

PR 05-MAY-1998; 98US-0072596.

PR 18-FEB-1998; 98US-0024753.

XX (CORI-) CORIXA CORP.

XX Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;

PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;

XX WPI; 1999-527416/44.

DR P-PSDB; AAY39058.

XX New polypeptide comprising antigenic portions of M. tuberculosis

PT Claim 11a; Page 274; 323pp; English.

XX This invention describes novel recombinant antigens and their encoding
CC nucleic acids derived from Mycobacterium tuberculosis. The novel
CC polypeptides are useful for detecting M. tuberculosis infection in a
CC biological sample by detecting antibodies which bind with the
CC polypeptides, and are useful as vaccines for immunizing against
CC M. tuberculosis infection. The new detection methods are needed as
CC current vaccination strategies do not provide 100% immunity.
XX

Sequence 423 BP; 95 A; 128 C; 129 G; 71 T; 0 other;

Query Match 28.8%; Score 15; DB 20; Length 423;

	Best Local Similarity	100.0%	Pred. No.	94						
	Matches	15	Conservative	0	Mismatches	0	Indels	0	Gaps	0
Qy	30	CGTCGCTGCCGGCCT	44							
Db	26	CGTCCCTGCCGGCCT	12							

RESULT 36	
ABQ62561	
ID	ABQ62561 standard; DNA; 448 BP.
XX	
AC	
XX	ABQ62561;
XX	
DT	16-AUG-2002 (first entry)
XX	
DE	Mycobacterium tuberculosis BAC vector clone RV129SP6.
XX	
KW	Mycobacterium tuberculosis; Mycobacterium bovis; mycobacterium;
KW	detection; BAC vector; bacterial artificial chromosome; tuberculosis;
XX	gene; ds.

OS Mycobacterium tuberculosis.
XX
XX
XX MO9954487-A2.
XX
XX
XX 28-OCT-1999.
XX
XX
XX 16-APR-1999; 99MO-IB00740.
XX
XX 16-APR-1998; 98US-0060756.
XX
XX (INSP) INST PASTEUR.
XX
XX
XX
XX Cole S, Buchrieser-Brosch R, Gordon S, Billault A;
XX WPI; 2000-013262/01.
XX

PT Isolating of polynucleotides from mycobacterial genomes, useful for
 PT detection of Mycobacteriia and for combating tuberculosis -
 XX
 XX
 PS Claim 23, Page 44-45, 161pp; English.
 CC
 CC The present invention describes a method for isolating a polynucleotide
 CC of interest that is present or is expressed in a genome of a first
 CC mycobacterium strain and that is absent or altered in a genome of a
 CC second mycobacterium strain, which is different from the first strain
 CC using a bacterial artificial chromosome (BAC) vector. Recombinant BAC
 CC vectors, which are preferably immobilised, can be used to detect
 CC mycobacterial nucleic acids (genomic DNA, cDNA or mRNA) in biological
 CC samples. The polynucleotides identified are useful as probes or primers
 CC for detecting a given mycobacterium of interest. By aligning the
 CC polynucleotides contained in the recombinant BAC vectors it is possible
 CC to physically map a polynucleotide of mycobacterial origin in a
 CC biological sample. The methods and vectors from the present invention
 CC are useful in providing information for combating tuberculosis. It is
 CC possible to compare genomes between different strains or species and
 CC their non-pathogenic strains or species counterparts. AB062492 to
 CC AB063228 and AB881227 to AB881230 represent sequences used in the
 CC exemplification of the present invention.

SQ Sequence 448 BP; 67 A; 141 C; 133 G; 102 T; 5 other;
 Query Match: 28.8%; Score 15; DB 21; Length 448;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

QY      7 TTCTCTCAGGCTCTC 21
          |||||
Db     352 TTCTCTCAGGCTCTC 366

```

RESULT 37

AAI90273
ID AAI90273 standard; cDNA; 459 BP

DT	06-NOV-2001	(first entry)
XX		
DE	Human polynucleotide SEQ ID NO 10333	

Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 tissue growth factor; immunomodulatory; cancer; leukaemia;
 nervous system disorders; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200164835-A2

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

X

3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842

XX
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700

DR P-PSDB; AA010342

PT Isolated nucleic acids and polypeptides, useful for preventing

PT disorders -

PS Claim 1; SEQ ID NO 103333; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haemopoietic regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.

Sequence 459 BP; 110 A; 142 C; 111 G; 95 T; 1 other;

Query Match	28.8%	Score 15;	DB 22;	Length 459;
-------------	-------	-----------	--------	-------------

```
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

17 CTCTCATCGCTCTCG 31

Db 199 CTCTCATCGCTCTCG 213

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

AAC02001
ID 11500001 4-1-1 5000 400 500

XXXXXX

XX
E
C
O
N
O
M
I
C

XX
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840

DE Human secreted protein 5' EST, SEQ ID NO: 1999.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX Homo sapiens.
 OS EPI033401-A2.
 PN
 XX 06-SEP-2000.
 PD
 XX 21-FEB-2000; 2000EP-0200610.
 PF
 XX 26-FEB-1999; 99US-0122487.
 PR
 XX (GEST) GENSET.
 PA
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI; 2000-500381/45.
 DR P-PSDB; AAG01995.
 XX
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PS Claim 1; SEQ ID 1999; 71pp + CD-ROM; English.
 XX
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.
 XX
 XX Sequence 471 BP; 142 A; 87 C; 111 G; 131 T; 0 other;
 SQ
 Query Match 28.8%; Score 15; DB 21; Length 471;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCAGTCTCTCTAC 15
 |||||
 Db 124 ATGCAGTCTCTCTAC 138
 RESULT 39
 ABA58638
 ID ABA58638 standard; DNA; 481 BP.
 XX
 AC ABA58638;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #6943.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX Homo sapiens.
 OS MO200157277-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00669.
 PF
 XX

PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-483447/52.
 DR
 XX
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 PS Claim 1; SEQ ID NO 6943; 639pp + sequence listing; English.
 XX
 XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 481 BP; 73 A; 136 C; 180 G; 92 T; 0 other;
 SQ
 Query Match 28.8%; Score 15; DB 22; Length 481;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 15 CGCTCTCATCGCTCT 29
 |||||
 Db 152 CGCTCTCATCGCTCT 166
 RESULT 40
 ABA27633
 ID ABA27633 standard; DNA; 481 BP.
 XX
 AC ABA27633;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Probe #6099 for gene expression analysis in human heart cell sample.
 XX
 KW Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 XX Homo sapiens.
 OS
 XX
 PN MO200157274-A2.
 XX
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00666.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI

```

XX MPI: 2001-488899/53.
DR
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID No 6099; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 481 BP; 73 A; 136 C; 180 G; 92 T; 0 other;

Query Match          28.8%; Score 15; DB 22; Length 481;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      15 CGCTCTCATCGCTCT 29
        |||||
        152 CGCTCTCATCGCTCT 166
  
```

Search completed: November 13, 2003, 09:52:46
 Job time : 249 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 09:44:31 / Search time 52 Seconds

(without alignments)
441.383 Million cell updates/sec

Title: US-10-081-935-1

Perfect score: 52
1 atgcagctctctcaagctctc.....cgctgcgcgcctcgcagctg 52

Scoring table:

Gapop_60.0, Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	38.5	732	US-09-252-991A-13272	Sequence 13272, A
2	20	38.5	1056	US-09-252-991A-13107	Sequence 13107, A
3	16	30.8	810	US-09-252-991A-13647	Sequence 13647, Ap
4	16	30.8	879	US-09-252-991A-13602	Sequence 13602, Ap
5	16	30.8	1632	US-09-252-991A-13859	Sequence 13859, Ap
6	16	30.8	2673	US-09-252-991A-13993	Sequence 13993, Ap
7	16	30.8	2715	US-09-252-991A-1686	Sequence 1686, Ap
8	16	30.8	2814	US-09-252-991A-1875	Sequence 1875, Ap
9	15	28.8	423	US-09-072-596-259	Sequence 259, App
10	15	28.8	438	US-09-252-991A-15918	Sequence 15918, A
11	15	28.8	448	US-09-060-756-60	Sequence 60, Appl
12	15	28.8	448	US-09-670-314-60	Sequence 60, Appl
13	15	28.8	762	US-09-252-991A-15888	Sequence 15888, A
14	15	28.8	883	US-09-252-991A-15774	Sequence 15774, A
15	15	28.8	1068	US-09-252-991A-15804	Sequence 15804, A
16	15	28.8	1119	US-09-252-991A-10788	Sequence 10788, A
17	15	28.8	1146	US-09-252-991A-10603	Sequence 10603, A
18	15	28.8	1197	US-09-620-312D-168	Sequence 168, App
19	15	28.8	1260	US-09-252-991A-10250	Sequence 10250, A
20	15	28.8	1371	US-09-252-991A-11745	Sequence 11745, A
21	15	28.8	1620	US-09-252-991A-10368	Sequence 10368, A
22	15	28.8	1692	US-09-252-991A-11887	Sequence 11887, A
23	15	28.8	2553	US-09-252-991A-1113	Sequence 1113, App
24	15	28.8	2664	US-09-252-991A-108	Sequence 108, App
25	15	28.8	4198	US-09-586-719-11	Sequence 11, Appl
26	15	28.8	4649	US-09-586-719-11	Patent No. 5183745
27	15	28.8	5118	US-08-669-785-3	Sequence 3, Appl1

28	15	28.8	6441	US-08-669-785-1	Sequence 1, Appl1
29	15	28.8	6443	5183745-5	Patent No. 5183745
30	15	28.8	31880	US-09-453-702B-242	Sequence 242, App
31	15	28.8	38653	US-09-922-445-1	Sequence 1, Appl1
32	15	28.8	44377	US-08-804-127C-7	Sequence 7, Appl1
33	15	28.8	44377	US-08-804-198-1	Sequence 1, Appl1
34	15	28.8	68750	US-09-335-409-1	Sequence 1, Appl1
35	15	28.8	68750	US-09-568-102-1	Sequence 1, Appl1
36	15	28.8	68750	US-09-567-969-1	Sequence 1, Appl1
37	15	28.8	68750	US-09-568-480-1	Sequence 1, Appl1
38	15	28.8	68750	US-09-568-486-1	Sequence 1, Appl1
39	15	28.8	68750	US-09-568-472-1	Sequence 1, Appl1
40	15	28.8	68750	US-09-567-899-1	Sequence 1, Appl1
41	15	28.8	71989	US-09-443-509A-1	Sequence 2, Appl1
42	15	28.8	536165	US-09-214-808-1	Sequence 1, Appl1
43	15	28.8	4403765	US-09-103-840A-2	Sequence 2, Appl1
44	15	28.8	4403765	US-09-103-840A-2	Sequence 2, Appl1
45	15	28.8	4411529	US-09-103-840A-1	Sequence 1, Appl1
46	15	28.8	4411529	US-09-103-840A-1	Sequence 1, Appl1
47	14	26.9	234	US-09-397-787-36	Sequence 36, Appl1
48	14	26.9	258	US-09-252-991A-1283	Sequence 1283, Ap
49	14	26.9	308	US-09-312-283C-110	Sequence 110, App
50	14	26.9	310	US-09-188-930-110	Sequence 110, App
51	14	26.9	405	US-09-252-991A-2315	Sequence 2315, App
52	14	26.9	417	US-09-252-991A-12658	Sequence 12658, A
53	14	26.9	486	US-09-252-991A-11068	Sequence 11068, A
54	14	26.9	603	US-09-252-991A-1063	Sequence 1063, Ap
55	14	26.9	651	US-09-252-991A-4491	Sequence 4491, Ap
56	14	26.9	792	US-09-252-991A-7005	Sequence 7005, Ap
57	14	26.9	813	US-09-252-991A-11103	Sequence 11103, A
58	14	26.9	846	US-09-252-991A-9635	Sequence 9635, Ap
59	14	26.9	927	US-09-252-991A-7558	Sequence 7558, Ap
60	14	26.9	939	US-09-252-991A-3216	Sequence 3216, Ap
61	14	26.9	960	US-08-651-136C-1	Sequence 1, Appl1
62	14	26.9	960	US-09-229-911A-1	Sequence 1, Appl1
63	14	26.9	1002	US-09-252-991A-1099	Sequence 1099, Ap
64	14	26.9	1008	US-09-252-991A-10996	Sequence 10996, A
65	14	26.9	1020	US-09-252-991A-5960	Sequence 5960, Ap
66	14	26.9	1062	US-09-252-991A-9532	Sequence 9532, Ap
67	14	26.9	1110	US-09-252-991A-9578	Sequence 9578, Ap
68	14	26.9	1119	US-09-252-991A-10559	Sequence 10559, A
69	14	26.9	1209	US-09-252-991A-5892	Sequence 5892, Ap
70	14	26.9	1230	US-09-252-991A-947	Sequence 947, App
71	14	26.9	1239	US-09-252-991A-6622	Sequence 6622, App
72	14	26.9	1275	US-09-252-991A-2399	Sequence 2399, Ap
73	14	26.9	1287	US-09-252-991A-11255	Sequence 11255, A
74	14	26.9	1371	US-09-252-991A-6010	Sequence 6010, Ap
75	14	26.9	1371	US-09-252-991A-9717	Sequence 9717, Ap
76	14	26.9	1410	US-08-343-101A-6	Sequence 6, Appl1
77	14	26.9	1410	US-09-183-688-6	Sequence 6, Appl1
78	14	26.9	1410	US-09-519-485-6	Sequence 6, Appl1
79	14	26.9	1425	US-09-252-991A-10301	Sequence 10301, A
80	14	26.9	1449	US-09-252-991A-6642	Sequence 6642, Ap
81	14	26.9	1521	US-09-252-991A-989	Sequence 989, App
82	14	26.9	1530	US-09-252-991A-9942	Sequence 9942, Ap
83	14	26.9	1548	US-09-252-991A-4907	Sequence 4907, Ap
84	14	26.9	1602	US-09-252-991A-4836	Sequence 4836, Ap
85	14	26.9	1620	US-09-252-991A-12984	Sequence 12984, A
86	14	26.9	1647	US-09-252-991A-5576	Sequence 5576, Ap
87	14	26.9	1653	US-09-252-991A-2515	Sequence 2515, Ap
88	14	26.9	1710	US-09-252-991A-6299	Sequence 6299, Ap
89	14	26.9	1731	US-09-252-991A-6575	Sequence 6575, Ap
90	14	26.9	1800	US-09-252-991A-12496	Sequence 12496, A
91	14	26.9	1857	US-09-322-478-24	Sequence 24, Appl1
92	14	26.9	1965	US-09-252-991A-4464	Sequence 4464, Ap
93	14	26.9	1986	US-09-252-991A-10499	Sequence 10499, Ap
94	14	26.9	2028	US-09-252-991A-7091	Sequence 7091, Ap
95	14	26.9	2235	US-09-252-991A-7032	Sequence 7032, Ap
96	14	26.9	2634	US-08-907-166-7	Sequence 7, Appl1
97	14	26.9	2634	US-09-391-340-7	Sequence 7, Appl1
98	14	26.9	2861	US-08-770-301A-12	Sequence 12, Appl1
99	14	26.9	2861	US-09-175-581-12	Sequence 12, Appl1
100	14	26.9	3105	US-09-252-991A-4398	Sequence 4398, Ap


```

c 977      12      23.1      1668      4      US-09-252-991A-12163      Sequence 12163, A
c 978      12      23.1      1674      4      US-09-252-991A-5894      Sequence 5894, Ap
c 979      12      23.1      1677      4      US-09-199-637A-276      Sequence 276, App
980      12      23.1      1677      4      US-09-252-991A-6747      Sequence 6747, Ap
981      12      23.1      1677      4      US-09-252-991A-7476      Sequence 7476, Ap
c 982      12      23.1      1683      4      US-09-252-991A-6977      Sequence 6977, Ap
c 983      12      23.1      1683      4      US-09-252-991A-11588      Sequence 11588, A
984      12      23.1      1692      4      US-09-252-991A-978      Sequence 978, App
c 985      12      23.1      1692      4      US-09-252-991A-14477      Sequence 14477, A
c 986      12      23.1      1698      4      US-09-252-991A-3551      Sequence 3551, Ap
c 987      12      23.1      1700      3      US-08-860-519-5      Sequence 5, Appl1
c 988      12      23.1      1700      3      US-08-860-519-6      Sequence 6, Appl1
c 989      12      23.1      1701      4      US-09-252-991A-7927      Sequence 7927, Ap
990      12      23.1      1707      4      US-09-207-388-10      Sequence 10, Appl
991      12      23.1      1710      4      US-09-252-991A-9057      Sequence 9057, Ap
992      12      23.1      1723      1      US-08-241-766-1      Sequence 1, Appl1
c 993      12      23.1      1723      1      US-08-241-766-2      Sequence 2, Appl1
c 994      12      23.1      1724      3      US-09-385-259-1      Sequence 1, Appl1
c 995      12      23.1      1724      3      US-09-645-370-1      Sequence 1, Appl1
c 996      12      23.1      1728      4      US-09-252-991A-616      Sequence 616, App
997      12      23.1      1734      4      US-09-252-991A-3996      Sequence 3996, Ap
998      12      23.1      1734      4      US-09-252-991A-9061      Sequence 9061, Ap
999      12      23.1      1734      4      US-09-252-991A-9061      Sequence 2, Appl1
1000      12      23.1      1736      4      US-09-162-524-2

```

ALIGNMENTS

```

RESULT 1
US-09-252-991A-13272/c
: Sequence 13272, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 13272
: LENGTH: 732
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13272

Query Match      38.5%; Score 20; DB 4; Length 732;
Best Local Similarity 100.0%; Pred. No. 0.068;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      28      CTCGTCGTCGCGGCGCTCGC 47
Db      381      CTCGTCGTCGCGGCGCTCGC 362

RESULT 2
US-09-252-991A-13107/c
: Sequence 13107, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 13107
: LENGTH: 732
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13107

Query Match      38.5%; Score 20; DB 4; Length 732;
Best Local Similarity 100.0%; Pred. No. 0.068;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 13107
: LENGTH: 1056
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13107

```

```

Query Match      38.5%; Score 20; DB 4; Length 1056;
Best Local Similarity 100.0%; Pred. No. 0.066;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      28      CTCGTCGTCGCGGCGCTCGC 47
Db      634      CTCGTCGTCGCGGCGCTCGC 615

```

```

RESULT 3
US-09-252-991A-3647/c
: Sequence 3647, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 3647
: LENGTH: 810
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3647

```

```

Query Match      30.8%; Score 16; DB 4; Length 810;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      33      CGCTGCCGCGGCGCTCGC 48
Db      286      CGCTGCCGCGGCGCTCGC 271

```

```

RESULT 4
US-09-252-991A-3602/c
: Sequence 3602, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 3602
: LENGTH: 879
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3602

```

```

Query Match      30.8%; Score 16; DB 4; Length 879;
Best Local Similarity 100.0%; Pred. No. 7.9;

```

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 33 CGCTGCCGCGCTGCC 48
Db 338 CGCTGCCGCGCTGCC 323

RESULT 5
US-09-252-991A-3859
; Sequence 3859, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3859
; LENGTH: 1632
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3859

Query Match 30.8%; Score 16; DB 4; Length 1632;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 33 CGCTGCCGCGCTGCC 48
Db 1353 CGCTGCCGCGCTGCC 1368

RESULT 6
US-09-252-991A-2993/C
; Sequence 2993, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2993
; LENGTH: 2673
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2993

Query Match 30.8%; Score 16; DB 4; Length 2673;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 GCTCTGCTGCTGCCG 40
Db 133 GCTCTGCTGCTGCCG 118

RESULT 7
US-09-252-991A-2686
; Sequence 2686, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2686
; LENGTH: 2715
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2686

Query Match 30.8%; Score 16; DB 4; Length 2715;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 GCTCTGCTGCTGCCG 40
Db 2652 GCTCTGCTGCTGCCG 2667

RESULT 8
US-09-252-991A-2875
; Sequence 2875, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2875
; LENGTH: 2814
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2875

Query Match 30.8%; Score 16; DB 4; Length 2814;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 GCTCTGCTGCTGCCG 40
Db 2710 GCTCTGCTGCTGCCG 2725

RESULT 9
US-09-072-596-259/C
; Sequence 259, Application US/09072596
; Patent No. 6458356
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Meto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-6031
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 259:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-072-596-259

Query Match 28.8%; Score 15; DB 4; Length 423;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 GCTCGCTCGCGCCT 44
|||||
Db 26 GCTCGCTCGCGCCT 12

RESULT 10
US-09-252-991A-15918
Sequence 15918, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15918
LENGTH: 438
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15918

Query Match 28.8%; Score 15; DB 4; Length 438;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 GCTGCGGCTTCGCC 48
|||||
Db 136 GCTGCGGCTTCGCC 150

RESULT 11
US-09-060-756-60
Sequence 60, Application US/09060756
Patent No. 6183957
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 60
LENGTH: 448
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: unsure
LOCATION: (various positions within the sequence)
OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-60

Query Match 28.8%; Score 15; DB 3; Length 448;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TTCTCTCAGCTCTC 21
|||||
Db 352 TTCTCTCAGCTCTC 366

RESULT 12
US-09-670-314-60
Sequence 60, Application US/09670314
Patent No. 6492506
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/670,314
CURRENT FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: 09/060,756
PRIOR FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 60
LENGTH: 448
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: unsure
LOCATION: (various positions within the sequence)
OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-670-314-60

Query Match 28.8%; Score 15; DB 4; Length 448;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TTCTCTCAGCTCTC 21
|||||
Db 352 TTCTCTCAGCTCTC 366

```
RESULT 13
US-09-252-991A-15888
; Sequence 15888, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15888
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15888

Query Match      28.8%; Score 15; DB 4; Length 762;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 GCTGCCGCGCTCGCC 48
DB      485 GCTGCCGCGCTCGCC 499

RESULT 14
US-09-252-991A-15774/C
; Sequence 15774, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15774
; LENGTH: 885
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15774

Query Match      28.8%; Score 15; DB 4; Length 885;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 GCTGCCGCGCTCGCC 48
DB      274 GCTGCCGCGCTCGCC 260

RESULT 15
US-09-252-991A-15804/C
; Sequence 15804, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
```

```
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15804
; LENGTH: 1068
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15804

Query Match      28.8%; Score 15; DB 4; Length 1068;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 GCTGCCGCGCTCGCC 48
DB      495 GCTGCCGCGCTCGCC 481

RESULT 16
US-09-252-991A-10788/C
; Sequence 10788, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10788
; LENGTH: 1119
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10788

Query Match      28.8%; Score 15; DB 4; Length 1119;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 GCTGCCGCGCTCGCC 48
DB      685 GCTGCCGCGCTCGCC 671

RESULT 17
US-09-252-991A-10603/C
; Sequence 10603, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10603
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10603
```


Query Match 28.8%; Score 15; DB 4; Length 1146;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 GCTGCCGCGCTCGCC 48
|||||
Db 735 GCTGCCGCGCTCGCC 721

RESULT 18

US-09-620-312D-168/C
; Sequence 168, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yuning
; APPLICANT: Wang, Dunxui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 168
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (96)...(548)
US-09-620-312D-168

Query Match 28.8%; Score 15; DB 4; Length 1197;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 CTCATCGCTCTGCTC 33
|||||
Db 100 CTCATCGCTCTGCTC 86

RESULT 19

US-09-252-991A-10250
; Sequence 10250, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10250
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10250

Query Match 28.8%; Score 15; DB 4; Length 1260;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 GCTGCCGCGCTCGCC 48
|||||
Db 406 GCTGCCGCGCTCGCC 420

RESULT 20

US-09-252-991A-11745/C
; Sequence 11745, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11745
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11745

Query Match 28.8%; Score 15; DB 4; Length 1317;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 TCGTCGCTGCCGCC 43
|||||
Db 629 TCGTCGCTGCCGCC 615

RESULT 21

US-09-252-991A-10368
; Sequence 10368, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10368
; LENGTH: 1620
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10368

Query Match 28.8%; Score 15; DB 4; Length 1620;

Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 GCTGCGGCGCTGCC 48
Db 1266 GCTGCGGCGCTGCC 1280

RESULT 22
US-09-252-991A-11887/c
; Sequence 11887, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11887
; LENGTH: 1692
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11887

Query Match 28.8%; Score 15; DB 4; Length 1692;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TCCTCGCTGCCGCC 43
Db 273 TCCTCGCTGCCGCC 259

RESULT 23
US-09-252-991A-113
; Sequence 113, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 113
; LENGTH: 2553
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-113

Query Match 28.8%; Score 15; DB 4; Length 2553;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 CGTCGCTGCCGCCCT 44
Db 105 CGTCGCTGCCGCCCT 119

RESULT 24
US-09-252-991A-108/c
; Sequence 108, Application US/09252991A

; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 108
; LENGTH: 2664
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-108

Query Match 28.8%; Score 15; DB 4; Length 2664;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 30 CGTCGCTGCCGCCCT 44
Db 2449 CGTCGCTGCCGCCCT 2435

RESULT 25
US-09-586-719-11
; Sequence 11, Application US/09586719
; Patent No. 6294345
; GENERAL INFORMATION:
; APPLICANT: Zheng, Feng
; APPLICANT: Levin, Joshua Z.
; APPLICANT: Bauer, Michael W.
; TITLE OF INVENTION: Herbicide Target Genes and Methods
; FILE REFERENCE: PB/5-10909A
; CURRENT FILING DATE: 2000-06-05
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 4198
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-586-719-11

Query Match 28.8%; Score 15; DB 3; Length 4198;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TCTCTCAGCCTCTCA 22
Db 958 TCTCTCAGCCTCTCA 972

RESULT 26
5183745-1/c
; Patent No. 5183745
; APPLICANT: DANCHIN, ANTOINE; GLASER, PHILIPPE; KRIN, EVELYN;
; BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES
; TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
; BIOLOGICAL USES
; NUMBER OF SEQUENCES: 13
; CURRENT FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1
; LENGTH: 4649
5183745-1

Query Match 28.8%; Score 15; DB 6; Length 4649;

Best Local Similarity 100.0%; Pred. No. 23;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GCTGCCGCGCTCGCC 48
|||||
Db 4350 GCTGCCGCGCTCGCC 4336

RESULT 27

US-08-669-785-3/c
; Sequence 3, Application US/08669785
; Patent No. 6309648
; GENERAL INFORMATION:
; APPLICANT: Betsou, Fotini
; APPLICANT: Betsou, Peter
; APPLICANT: Guiso, Nicole
; TITLE OF INVENTION: Protective Epitopes Of Adenyl
; TITLE OF INVENTION: Cyclase-Haemolysin(AC-Hly), Their Application To
; TITLE OF INVENTION: The Treatment Or To The Prevention Of Bordetella Infections.
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,785
; FILING DATE: 27-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 02356.0072-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5118 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..5115
; OTHER INFORMATION: /note= "AMINO ACID SEQUENCE
; OTHER INFORMATION: CORRESPONDING TO THE NUCLEOTIDE SEQUENCE OF THE GENE
; OTHER INFORMATION: CODING FOR THE B. Bronchiseptica AC-Hly"
US-08-669-785-3
Query Match 28.8%; Score 15; DB 4; Length 5118;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 34 GCTGCCGCGCTCGCC 48
|||||
Db 4164 GCTGCCGCGCTCGCC 4150

RESULT 28

US-08-669-785-1/c
; Sequence 1, Application US/08669785
; Patent No. 6309648
; GENERAL INFORMATION:
; APPLICANT: Betsou, Fotini

; APPLICANT: Sebo, Peter
; APPLICANT: Guiso, Nicole
; TITLE OF INVENTION: Protective Epitopes Of Adenyl
; TITLE OF INVENTION: Cyclase-Haemolysin(AC-Hly), Their Application To
; TITLE OF INVENTION: The Treatment Or To The Prevention Of Bordetella Infections.
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,785
; FILING DATE: 27-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 02356.0072-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6441 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 979..6096
; OTHER INFORMATION: /note= "AMINO ACID SEQUENCE
; OTHER INFORMATION: CORRESPONDING TO THE NUCLEOTIDE SEQUENCE OF THE GENE
; OTHER INFORMATION: CODING FOR THE B. Pertussis AC-Hly"
US-08-669-785-1

Query Match 28.8%; Score 15; DB 4; Length 6441;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GCTGCCGCGCTCGCC 48
|||||
Db 5145 GCTGCCGCGCTCGCC 5131

RESULT 29
5183745-5/c
; Patent No. 5183745
; APPLICANT: DANCHIN, ANTOINE; GLASER, PHILIPPE; KRIN, EVELYN;
; BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES
; TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
; BIOLOGICAL USES
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,541
; FILING DATE: 25-OCT-1989
; SEQ ID NO: 5:
; LENGTH: 6443
5183745-5

Query Match 28.8%; Score 15; DB 6; Length 6443;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GCTGCCGCGCTCGCC 48

```
Db          5147 GCTGCCGCTGCC 5133

RESULT 30
US-09-453-702B-242/C
; Sequence 242, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blatner, Frederick R.
;           Burland, Valerie
;           Perna, Nicole T.
;           Plunkett, Guy
;           Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plunkney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296,95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 242:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31880
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 242:
US-09-453-702B-242

Query Match      28.8%; Score 15; DB 4; Length 31880;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      23 TCGCTCTGTCGCTG 37
Db      26165 TCGCTCTGTCGCTG 26151

RESULT 31
US-09-922-445-1
; Sequence 1, Application US/09922445
; Patent No. 6528268
; GENERAL INFORMATION:
; APPLICANT: Andersson, Maria K.
;           Berglund, Lars G. T.
;           Berglund, Rikard H.
;           Adam, Gail I. R.
; TITLE OF INVENTION: REAGENTS AND METHODS FOR DETECTION OF HEART FAILURE
; FILE REFERENCE: GGI2605
; CURRENT APPLICATION NUMBER: US/09/922,445

; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 38653
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(26156)
; OTHER INFORMATION:
; NAME/KEY: misc_feature
; LOCATION: (24801)..(24801)
; OTHER INFORMATION: nucleotide 24801 is a single nucleotide polymorphism which can t
; NAME/KEY: misc_feature
; LOCATION: (24941)..(24941)
; OTHER INFORMATION: nucleotide 24941 is a single nucleotide polymorphism which can t
; OTHER INFORMATION: T or C
; NAME/KEY: exon
; LOCATION: (26157)..(26252)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (26253)..(26401)
; OTHER INFORMATION:
; NAME/KEY: exon
; LOCATION: (26402)..(26543)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (26544)..(27024)
; OTHER INFORMATION:
; NAME/KEY: exon
; LOCATION: (27025)..(27178)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (27179)..(30519)
; OTHER INFORMATION:
; NAME/KEY: misc_feature
; LOCATION: (27645)..(27645)
; OTHER INFORMATION: nucleotide 27645 is a single nucleotide polymorphism which can t
; OTHER INFORMATION: C or G
; NAME/KEY: exon
; LOCATION: (30520)..(30681)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (30682)..(30894)
; OTHER INFORMATION:
; NAME/KEY: exon
; LOCATION: (30895)..(31027)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (31028)..(31747)
; OTHER INFORMATION:
; NAME/KEY: exon
; LOCATION: (31748)..(31841)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (31842)..(32400)
; OTHER INFORMATION:
; NAME/KEY: misc_feature
; LOCATION: (32163)..(32163)
; OTHER INFORMATION: nucleotide 32163 is a single nucleotide polymorphism which can t
; OTHER INFORMATION: A or C
; NAME/KEY: exon
; LOCATION: (32401)..(32528)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (32529)..(33414)
; OTHER INFORMATION:
; NAME/KEY: misc_feature
; LOCATION: (32614)..(32614)
; OTHER INFORMATION: nucleotide 32614 is a single nucleotide polymorphism which can b
; OTHER INFORMATION: A or G
```

```

? NAME/KEY: exon
? LOCATION: (33415)..(33597)
? OTHER INFORMATION:
? NAME/KEY: Intron
? LOCATION: (33598)..(34314)
? OTHER INFORMATION:
? NAME/KEY: exon
? LOCATION: (34315)..(34588)
? OTHER INFORMATION:
? NAME/KEY: Intron
? LOCATION: (34589)..(36404)
? OTHER INFORMATION:
? NAME/KEY: exon
? LOCATION: (36405)..(36523)
? OTHER INFORMATION:
? NAME/KEY: Intron
? LOCATION: (36524)..(38341)
? OTHER INFORMATION:
? NAME/KEY: exon
? LOCATION: (38342)..(38653)
? OTHER INFORMATION:
? PUBLICATION INFORMATION:
? DATABASE ACCESSION NUMBER: Genbank/AC004923
? DATABASE ENTRY DATE: 1999-12-21
? RELEVANT RESIDUES: (1)..(38653)
US-09-922-445-1
```

```
Query Match      28.8%; Score 15; DB 4; Length 38653;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      15 CGCTTCATCGCTCT 29
Db      8343 CGCTTCATCGCTCT 8357
```

```

RESULT 32
US-08-804-227C-7
? Sequence 7, Application US/08804227C
? Patent No. 5876991
? GENERAL INFORMATION:
? APPLICANT: Dehoff, Bradley S.
? APPLICANT: Kuhstoss, Stuart A.
? APPLICANT: Rosteck, Paul R., Jr.
? APPLICANT: Sutton, Kimberly L.
? TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
? NUMBER OF SEQUENCES: 15
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: THOMAS G. PLANT 1501
? STREET: LILLY CORPORATE CENTER
? CITY: INDIANAPOLIS
? STATE: IN
? COUNTRY: USA
? ZIP: 46285
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: ASCII(DOS) Text only
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/804,227C
? FILING DATE: February 21, 1997
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Plant, Thomas G.
? REGISTRATION NUMBER: 35,784
? REFERENCE/DOCKET NUMBER: X-8231
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 317-276-2459
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 44377 base pairs
? TYPE: nucleic acid
```

```

? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 350..14002
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 14046..20036
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 20110..31284
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 31329..36071
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 36155..41830
US-08-804-227C-7
```

```
Query Match      28.8%; Score 15; DB 2; Length 44377;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      34 GCTGCCGCGCTCGCC 48
Db      17373 GCTGCCGCGCTCGCC 17387
```

```

RESULT 33
US-08-804-198-1
? Sequence 1, Application US/08804198
? Patent No. 5945320
? GENERAL INFORMATION:
? APPLICANT: Burgett, Stanley G.
? APPLICANT: Kuhstoss, Stuart A.
? APPLICANT: Rao, Nagaraja R.
? APPLICANT: Richardson, Mark A.
? APPLICANT: Rosteck, Paul R., Jr.
? TITLE OF INVENTION: PLATENOILIDE SYNTHASE GENE
? NUMBER OF SEQUENCES: 6
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: PAUL R. CANTRELL 1138
? STREET: LILLY CORPORATE CENTER
? CITY: INDIANAPOLIS
? STATE: IN
? COUNTRY: USA
? ZIP: 46285
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: Macintosh
? SOFTWARE: Microsoft Word 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/804,198
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: CANTRELL, PAUL R.
? REGISTRATION NUMBER: 36,470
? REFERENCE/DOCKET NUMBER: P9113
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 317-276-3885
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 44377 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 350..14002
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-198-1
```

```
Query Match          28.8%; Score 15; DB 2; Length 44377;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      34 GCTGCCGCGCTCGCC 48
          |||||
Db      17373 GCTGCCGCGCTCGCC 17387
```

```
RESULT 34
US-09-335-409-1/c
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
```

```
Query Match          28.8%; Score 15; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      32 TCGCTGCCGCGCTCG 46
          |||||
Db      6467 TCGCTGCCGCGCTCG 6453
```

```
RESULT 35
US-09-568-102-1/c
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
```

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
```

```
Query Match          28.8%; Score 15; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      32 TCGCTGCCGCGCTCG 46
          |||||
Db      6467 TCGCTGCCGCGCTCG 6453
```

```
RESULT 36
US-09-567-969-1/c
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-969-1
```

```
Query Match          28.8%; Score 15; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      32 TCGCTGCCGCGCTCG 46
          |||||
Db      6467 TCGCTGCCGCGCTCG 6453
```

```
RESULT 37
US-09-568-480-1/c
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
```

US-09-568-480-1

Query Match 28.8%; Score 15; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TCGCTGCCGCGCCTCG 46

DB 6467 TCGCTGCCGCGCCTCG 6453

RESULT 38

US-09-568-486-1/c
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Query Match 28.8%; Score 15; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TCGCTGCCGCGCCTCG 46

DB 6467 TCGCTGCCGCGCCTCG 6453

RESULT 39

US-09-568-472-1/c
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match 28.8%; Score 15; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TCGCTGCCGCGCCTCG 46
DB 6467 TCGCTGCCGCGCCTCG 6453

RESULT 40

US-09-567-899-1/c
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-567-899-1

Query Match 28.8%; Score 15; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TCGCTGCCGCGCCTCG 46

DB 6467 TCGCTGCCGCGCCTCG 6453

Search completed: November 13, 2003, 10:59:54
Job time : 79 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 10:26:42 : Search time 214 Seconds

(without alignments)
794.143 Million cell updates/sec

Title: US-10-081-935-1

Perfect score: 52

Sequence: 1 atgcagctctctcagcctc.....cgctcgcgagcctcgccagtg 52

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 2169961 seqs, 1634102185 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications, NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17	32.7	1392	12	US-10-286-606-11
C 2	17	32.7	1392	14	US-10-126-279-11
C 3	16	30.8	574	11	US-09-918-995-25977
C 4	16	30.8	1653	9	US-09-897-214-7
C 5	16	30.8	1872	9	US-09-768-826-21
C 6	16	30.8	2451	12	US-10-093-463-189
C 7	16	30.8	2563	12	US-10-093-463-191
C 8	16	30.8	2610	12	US-10-246-310-11
C 9	16	30.8	3701	12	US-10-923-463-187
C 10	16	30.8	3947	10	US-09-925-300-638
C 11	16	30.8	3953	14	US-10-106-698-1931
C 12	15	28.8	214	9	US-09-294-093B-4207
C 13	15	28.8	264	10	US-09-878-574-15763
C 14	15	28.8	294	14	US-10-156-761-711
C 15	15	28.8	336	10	US-09-878-574-1259
C 16	15	28.8	357	9	US-09-764-870-129

C 17	15	28.8	357	14	US-10-125-540-129	Sequence 129, App
C 18	15	28.8	423	12	US-10-084-884-254	Sequence 264, App
C 19	15	28.8	423	12	US-10-193-002-259	Sequence 259, App
C 20	15	28.8	481	12	US-10-259-678-60	Sequence 60, App
C 21	15	28.8	481	9	US-09-864-761-6099	Sequence 6099, App
C 22	15	28.8	603	12	US-10-027-632-115584	Sequence 115584, App
C 23	15	28.8	603	13	US-10-027-632-115584	Sequence 115584, App
C 24	15	28.8	1113	14	US-10-156-761-974	Sequence 974, App
C 25	15	28.8	1191	12	US-10-259-165-279	Sequence 279, App
C 26	15	28.8	1197	14	US-10-037-270-168	Sequence 168, App
C 27	15	28.8	1389	14	US-10-156-761-780	Sequence 780, App
C 28	15	28.8	1404	14	US-10-177-293-416	Sequence 416, App
C 29	15	28.8	1479	14	US-10-156-761-4461	Sequence 4461, App
C 30	15	28.8	1554	14	US-10-128-714-7070	Sequence 7070, App
C 31	15	28.8	1778	14	US-10-128-714-6070	Sequence 6070, App
C 32	15	28.8	1932	10	US-09-801-368-235	Sequence 235, App
C 33	15	28.8	2000	10	US-09-938-842A-2978	Sequence 2978, App
C 34	15	28.8	2412	14	US-10-156-761-158	Sequence 158, App
C 35	15	28.8	2454	9	US-09-815-242-4138	Sequence 4138, App
C 36	15	28.8	2672	11	US-09-374-046A-125	Sequence 125, App
C 37	15	28.8	2790	10	US-09-954-456-518	Sequence 518, App
C 38	15	28.8	2790	10	US-09-967-768A-182	Sequence 182, App
C 39	15	28.8	3140	14	US-10-128-714-70	Sequence 70, App
C 40	15	28.8	3323	11	US-09-822-846-197	Sequence 197, App
C 41	15	28.8	3778	14	US-10-128-714-5070	Sequence 5070, App
C 42	15	28.8	3963	12	US-10-161-051-168	Sequence 168, App
C 43	15	28.8	4463	12	US-10-214-350-3	Sequence 3, App
C 44	15	28.8	6059	12	US-10-240-452-43	Sequence 43, App
C 45	15	28.8	6442	10	US-09-950-335A-11	Sequence 11, App
C 46	15	28.8	10431	14	US-10-198-846-11715	Sequence 11715, App
C 47	15	28.8	10438	12	US-09-814-351-19864	Sequence 19864, App
C 48	15	28.8	11847	14	US-10-270-333-187	Sequence 187, App
C 49	15	28.8	31880	13	US-10-114-170-242	Sequence 242, App
C 50	15	28.8	68750	14	US-10-014-1717-1	Sequence 1, App
C 51	15	28.8	536165	14	US-09-939-964-1	Sequence 1, App
C 52	15	28.8	9026608	14	US-10-156-761-1	Sequence 1, App
C 53	15	28.8	9026608	14	US-10-156-761-1	Sequence 1, App
C 54	14	26.9	209	9	US-09-823-876-2874	Sequence 2874, App
C 55	14	26.9	234	9	US-09-876-889-36	Sequence 36, App
C 56	14	26.9	255	14	US-10-156-761-1349	Sequence 1349, App
C 57	14	26.9	310	11	US-09-866-050A-110	Sequence 110, App
C 58	14	26.9	310	14	US-10-152-661-1107	Sequence 1107, App
C 59	14	26.9	384	10	US-09-878-574-1507	Sequence 1507, App
C 60	14	26.9	420	14	US-10-156-761-1546	Sequence 1546, App
C 61	14	26.9	427	9	US-09-912-020-64	Sequence 64, App
C 62	14	26.9	432	13	US-10-032-159A-5	Sequence 5, App
C 63	14	26.9	468	11	US-09-918-995-24558	Sequence 24558, App
C 64	14	26.9	492	14	US-10-156-761-1986	Sequence 1986, App
C 65	14	26.9	498	12	US-10-027-632-180608	Sequence 180608, App
C 66	14	26.9	498	12	US-10-027-632-180608	Sequence 180608, App
C 67	14	26.9	498	13	US-10-027-632-180608	Sequence 180608, App
C 68	14	26.9	498	13	US-10-027-632-180608	Sequence 180608, App
C 69	14	26.9	515	12	US-10-027-632-101390	Sequence 101390, App
C 70	14	26.9	516	12	US-10-027-632-16803	Sequence 16803, App
C 71	14	26.9	516	12	US-10-027-632-80036	Sequence 80036, App
C 72	14	26.9	516	13	US-10-027-632-36803	Sequence 36803, App
C 73	14	26.9	516	13	US-10-027-632-36803	Sequence 36803, App
C 74	14	26.9	522	9	US-10-027-632-80036	Sequence 80036, App
C 75	14	26.9	522	9	US-09-764-870-572	Sequence 572, App
C 76	14	26.9	522	9	US-09-764-870-572	Sequence 572, App
C 77	14	26.9	522	14	US-10-125-540-572	Sequence 572, App
C 78	14	26.9	522	14	US-10-125-540-573	Sequence 573, App
C 79	14	26.9	524	12	US-10-027-632-273527	Sequence 273527, App
C 80	14	26.9	524	12	US-10-027-632-273528	Sequence 273528, App
C 81	14	26.9	524	13	US-10-027-632-273527	Sequence 273527, App
C 82	14	26.9	524	13	US-10-027-632-273528	Sequence 273528, App
C 83	14	26.9	526	12	US-10-027-632-273529	Sequence 273529, App
C 84	14	26.9	526	13	US-10-027-632-273529	Sequence 273529, App
C 85	14	26.9	539	12	US-10-027-632-275107	Sequence 275107, App
C 86	14	26.9	539	13	US-10-027-632-275107	Sequence 275107, App
C 87	14	26.9	567	12	US-10-259-165-371	Sequence 371, App
C 88	14	26.9	573	12	US-10-259-165-371	Sequence 371, App
C 89	14	26.9	612	10	US-10-259-165-371	Sequence 371, App
C 90	14	26.9	612	10	US-09-938-842A-129	Sequence 129, App


```
C 966 13 25.0 1547 14 US-10-127-824A-417 Sequence 417, App
C 967 13 25.0 1547 14 US-10-127-826A-417 Sequence 417, App
C 968 13 25.0 1547 14 US-10-127-827A-417 Sequence 417, App
C 969 13 25.0 1547 14 US-10-127-828A-417 Sequence 417, App
C 970 13 25.0 1547 14 US-10-127-830A-417 Sequence 417, App
C 971 13 25.0 1547 14 US-10-127-832A-417 Sequence 417, App
C 972 13 25.0 1547 14 US-10-127-833A-417 Sequence 417, App
C 973 13 25.0 1547 14 US-10-127-834A-417 Sequence 417, App
C 974 13 25.0 1547 14 US-10-127-836A-417 Sequence 417, App
C 975 13 25.0 1547 14 US-10-127-841A-417 Sequence 417, App
C 976 13 25.0 1547 14 US-10-127-844A-417 Sequence 417, App
C 977 13 25.0 1547 14 US-10-128-687A-417 Sequence 417, App
C 978 13 25.0 1547 14 US-10-128-688A-417 Sequence 417, App
C 979 13 25.0 1547 14 US-10-128-689A-417 Sequence 417, App
C 980 13 25.0 1547 14 US-10-128-694A-417 Sequence 417, App
C 981 13 25.0 1547 14 US-10-131-825A-417 Sequence 417, App
C 982 13 25.0 1547 14 US-10-131-825A-417 Sequence 417, App
C 983 13 25.0 1547 14 US-10-131-815A-417 Sequence 417, App
C 984 13 25.0 1547 14 US-10-131-817A-417 Sequence 417, App
C 985 13 25.0 1547 14 US-10-131-821A-417 Sequence 417, App
C 986 13 25.0 1547 14 US-10-131-822A-417 Sequence 417, App
C 987 13 25.0 1547 14 US-10-131-828A-417 Sequence 417, App
C 988 13 25.0 1547 14 US-10-131-835A-417 Sequence 417, App
C 989 13 25.0 1547 14 US-10-137-864A-417 Sequence 417, App
C 990 13 25.0 1547 14 US-10-137-869A-417 Sequence 417, App
C 991 13 25.0 1547 14 US-10-147-523-417 Sequence 417, App
C 992 13 25.0 1547 14 US-10-158-785-417 Sequence 417, App
C 993 13 25.0 1547 14 US-10-121-051-417 Sequence 417, App
C 994 13 25.0 1547 14 US-10-121-042-417 Sequence 417, App
C 995 13 25.0 1547 14 US-10-123-912-417 Sequence 417, App
C 996 13 25.0 1547 14 US-10-166-709A-465 Sequence 465, App
C 997 13 25.0 1547 14 US-10-192-007-417 Sequence 417, App
C 998 13 25.0 1547 14 US-10-194-359-417 Sequence 417, App
C 999 13 25.0 1547 14 US-10-127-847A-417 Sequence 417, App
C1000 13 25.0 1547 14 US-10-175-590-417 Sequence 417, App
```

ALIGNMENTS

```
RESULT 1
US-10-286-606-11/C
; Sequence 11, Application US/10286606
; Publication No. US20030180321A1
; GENERAL INFORMATION:
; APPLICANT: Bertozzi, Carolyn
; APPLICANT: Williams, Spencer J.
; APPLICANT: Mougous, Joseph
; TITLE OF INVENTION: Mycobacterial Sulfation Pathway Proteins
; TITLE OF INVENTION: and Methods of Use Thereof
; FILE REFERENCE: BERK-012CIP
; CURRENT APPLICATION NUMBER: US/10/286,606
; PRIOR FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 60/285,394
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/345,953
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: 10/126,279
; PRIOR FILING DATE: 2002-04-19
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1392
; TYPE: DNA
; ORGANISM: Mycobacterium avium
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1392)
US-10-286-606-11

Query Match 32.7%; Score 17; DB 12; Length 1392;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 29 TCCTCGCTGCCGCGCTC 45
|||||
Db 290 TCCTCGCTGCCGCGCTC 274
```

```
RESULT 2
US-10-126-279-11/C
; Sequence 11, Application US/10126279
; Publication No. US20030104001A1
; GENERAL INFORMATION:
; APPLICANT: Bertozzi, Carolyn
; APPLICANT: Williams, Spencer J.
; APPLICANT: Mougous, Joseph
; TITLE OF INVENTION: Mycobacterial Sulfation Pathway Proteins
; TITLE OF INVENTION: and Methods of Use Thereof
; FILE REFERENCE: BERK-012
; CURRENT APPLICATION NUMBER: US/10/126,279
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/285,394
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/345,953
; PRIOR FILING DATE: 2001-10-26
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1392
; TYPE: DNA
; ORGANISM: Mycobacterium avium
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1392)
US-10-126-279-11
```

```
Query Match 32.7%; Score 17; DB 14; Length 1392;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 29 TCCTCGCTGCCGCGCTC 45
|||||
Db 290 TCCTCGCTGCCGCGCTC 274
```

```
RESULT 3
US-09-918-995-25977
; Sequence 25977, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25977
; LENGTH: 574
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(574)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-25977
```

```
Query Match 30.8%; Score 16; DB 11; Length 574;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 CTCATCGCTCTCTCTCG 34
```

```
Db          194 CTCATCGCTCTCCTCG 209
|||||
RESULT 4
US-09-897-214-7
; Sequence 7, Application US/09897214
; Patent No. US20020076779A1
; GENERAL INFORMATION:
; APPLICANT: Thayer, Edward C.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: Leucine-Rich Repeat Proteins, Zlr7,
; FILE REFERENCE: 01-27
; CURRENT APPLICATION NUMBER: US/09/897,214
; PRIORITY FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/215,446
; PRIORITY FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1653
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1653)
US-09-897-214-7

Query Match          30.8%; Score 16; DB 9; Length 1653;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          33 CGCTGCGGCGCTCGCC 48
|||||
Db          1496 CGCTGCGGCGCTCGCC 1511
|||||

RESULT 5
US-09-768-826-21
; Sequence 21, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; FILE REFERENCE: PFS12P1
; TITLE OF INVENTION: 18 human secreted proteins
; CURRENT APPLICATION NUMBER: US/09/768,826
; PRIORITY FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIORITY FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIORITY FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 1872
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-768-826-21

Query Match          30.8%; Score 16; DB 9; Length 1872;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          33 CGCTGCGGCGCTCGCC 48
|||||
Db          1153 CGCTGCGGCGCTCGCC 1168
|||||

RESULT 6
US-10-093-463-189
; Sequence 189, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glenda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Baha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyanekar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Paturnajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1 Antbodies that Bind to Antigenic Polypepi
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIORITY FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIORITY FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIORITY FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIORITY FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIORITY FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIORITY FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIORITY FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIORITY FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIORITY FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIORITY FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIORITY FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIORITY FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIORITY FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIORITY FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIORITY FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIORITY FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIORITY FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIORITY FILING DATE: 2001-03-13
```



```
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 189
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (471)..(2169)
US-10-093-463-189

Query Match          30.8%; Score 16; DB 12; Length 2451;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      33  CGCTGCGCGCTCGCC 48
          |||
          1966 CGCTGCGCGCTCGCC 1981

RESULT 7
US-10-093-463-191
; Sequence 191, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glenda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Paturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1 Antibodies that Bind to Antigenic Polypeptide
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; PRIOR FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
```

```
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 191
; LENGTH: 2563
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (130)..(2464)
US-10-093-463-191

Query Match          30.8%; Score 16; DB 12; Length 2563;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      33  CGCTGCGCGCTCGCC 48
          |||
          1679 CGCTGCGCGCTCGCC 1694

RESULT 8
US-10-246-330-1
; Sequence 1, Application US/10246330
; Publication No. US20030166030A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Pah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; FILE REFERENCE: 14537-002001
; CURRENT APPLICATION NUMBER: US/10/246,330
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/323,241
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2610
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2607)
US-10-246-330-1

Query Match          30.8%; Score 16; DB 12; Length 2610;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      25  GCTCTGCTGCTCGCCG 40
```

Db 2547 GCCTCGCTCGCTCGCC 2562

RESULT 9

US-10-093-463-187
; Sequence 187, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Khekda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Wezes, Peter
; APPLICANT: Smithson, Glenda
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Baha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malvankar, Uriel
; APPLICANT: Anderson, David
; APPLICANT: Patlurajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J., Jr.
; TITLE OF INVENTION: No. US20030208039A1 Antibodies that Bind to Antigenic Polypeptide
; FILE REFERENCE: 21402-290A (Cura 590AT)
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380

; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 187
; LENGTH: 3201
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (471)..(2412)
US-10-093-463-187

Query Match

Best Local Similarity 30.8%; Score 16; DB 12; Length 3201;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 33 CGCTGCGCGCTCGCC 48

Db 1966 CGCTGCGCGCTCGCC 1981

RESULT 10

US-09-925-300-638/c
; Sequence 638, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 638
; LENGTH: 3947
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (625)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (3738)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-638

Query Match

Best Local Similarity 30.8%; Score 16; DB 10; Length 3947;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 CTCATCGCTCTCGTCG 34

Db 72 CTCATCGCTCTCGTCG 57

RESULT 11

US-10-106-698-1931/c
; Sequence 1931, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide

```
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1931
; LENGTH: 3953
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (618)..(618)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc.feature
; LOCATION: (626)..(626)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1931
```

```
Query Match          30.8%; Score 16; DB 14; Length 3953;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      19 CTCATCGCTCTCGTCG 34
          |||||
Db       72 CTCATCGCTCTCGTCG 57
```

```
RESULT 12
US-09-294-093B-4207
; Sequence 4207, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 4207
; LENGTH: 214
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700353417H1
; NAME/KEY: unsure
; LOCATION: 10, 37, 204
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-4207
```

```
Query Match          28.8%; Score 15; DB 9; Length 214;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 ATCAGTTCTCTCAG 15
          |||||
Db      182 ATCAGTTCTCTCAG 196
```

```
RESULT 13
US-09-878-574-15763/c
; Sequence 15763, Application US/09878574
```

```
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 15763
; LENGTH: 264
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701070483H1
US-09-878-574-15763
```

```
Query Match          28.8%; Score 15; DB 10; Length 264;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      26 CTCCTCGCTGCGCG 40
          |||||
Db      111 CTCCTCGCTGCGCG 97
```

```
RESULT 14
US-10-156-761-711/c
; Sequence 711, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 711
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(294)
US-10-156-761-711
```

```
Query Match          28.8%; Score 15; DB 14; Length 294;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      29 TCGTGGCTGCGCGC 43
          |||||
Db      143 TCGTGGCTGCGCGC 129
```

```
RESULT 15
US-09-878-574-1259/c
; Sequence 1259, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
```

APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO: 1259
LENGTH: 336
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: LIB3028-040-Q1-B1-A1
US-09-878-574-1259

Query Match 28.8%; Score 15; DB 10; Length 336;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CTCACGCTCTCAT 23
DB 67 CTCACGCTCTCAT 53

RESULT 16
US-09-764-870-129
Sequence 129, Application US/09764870
Patent No. US20020042386A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT214
CURRENT APPLICATION NUMBER: US/09/764,870
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 646
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO: 129
LENGTH: 357
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-870-129

Query Match 28.8%; Score 15; DB 9; Length 357;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCACTTCTCTCAC 15
DB 102 ATGCACTTCTCTCAC 116

RESULT 17
US-10-125-540-129
Sequence 129, Application US/10125540
Publication No. US20030059875A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT214CT
CURRENT APPLICATION NUMBER: US/10/125,540
CURRENT FILING DATE: 2002-04-19
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 646
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO: 129
LENGTH: 357
TYPE: DNA
ORGANISM: Homo sapiens
US-10-125-540-129

Query Match 28.8%; Score 15; DB 14; Length 357;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCACTTCTCTCAC 15
DB 102 ATGCACTTCTCTCAC 116

RESULT 18
US-10-084-843-264/C
Sequence 264, Application US/10084843
Publication No. US20030143243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, David C.
Campos-Neco, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 264:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 264:
US-10-084-843-264

Query Match 28.8%; Score 15; DB 12; Length 423;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 GGTGCTGCGGCGCT 44
DB 26 GGTGCTGCGGCGCT 12

RESULT 19

US-10-193-002-259/c
; Sequence 259, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campos-Neto, Antonia
; Houghton, Raymond
; Vedvick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 259:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 259:
US-10-193-002-259
Query Match 28.8%; Score 15; DB 12; Length 423;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 30 CGTGGCTGCGGCGCT 44
DB 26 CGTGGCTGCGGCGCT 12
RESULT 20
US-10-259-678-60
; Sequence 60, Application US/10259678
; Publication No. US20030198974A1
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA

FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/10/259,678
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 60
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-10-259-678-60
Query Match 28.8%; Score 15; DB 12; Length 448;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 TTCTCAGCGCTCTC 21
DB 352 TTCTCAGCGCTCTC 366
RESULT 21
US-09-864-761-6099
; Sequence 6099, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408

```

; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6099
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004923.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 14
US-09-864-761-6099
```

```

Query Match          28.8%; Score 15; DB 9; Length 481;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 15 CGCTCTCATGCTCT 29
Db 152 CGCTCTCATGCTCT 166
```

```

RESULT 22
US-10-027-632-115584
; Sequence 115584, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115584
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-115584
```

```

Query Match          28.8%; Score 15; DB 12; Length 603;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 25 GCTCTGTCGCTGCC 39
Db 48 GCTCTGTCGCTGCC 62
```

RESULT 23

```

US-10-027-632-115584
; Sequence 115584, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115584
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-115584
```

```

Query Match          28.8%; Score 15; DB 13; Length 603;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 25 GCTCTGTCGCTGCC 39
Db 48 GCTCTGTCGCTGCC 62
```

```

RESULT 24
US-10-156-761-974/c
; Sequence 974, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 974
; LENGTH: 1113
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1113)
US-10-156-761-974
```

```

Query Match          28.8%; Score 15; DB 14; Length 1113;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 29 TCGTCGCTGCCGCC 43
|||||
Db 845 TCGTCGCTGCCGCC 831

RESULT 25

US-10-259-165-279
; Sequence 279, Application US/10259165
; Publication No. US2003013588A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kieps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Rieke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: Patentlist.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 279
; LENGTH: 1191
; TYPE: DNA
; ORGANISM: *Oryza sativa*
US-10-259-165-279

Query Match 28.8%; Score 15; DB 12; Length 1191;
Best Local Similarity 100.0%; Pred. No. 63;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 TCGCTGCCGCCCTCG 46
|||||
Db 88 TCGCTGCCGCCCTCG 102

RESULT 26

US-10-037-270-168/c
; Sequence 168, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and

; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_fl_genes Version 1.0
; SEQ ID NO 168
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: *Homo sapiens*
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (96)..(548)
US-10-037-270-168

Query Match 28.8%; Score 15; DB 14; Length 1197;
Best Local Similarity 100.0%; Pred. No. 63;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 CTCATCGCTCTGTC 33
|||||
Db 100 CTCATCGCTCTGTC 86

RESULT 27

US-10-156-761-780/c
; Sequence 780, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 780
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: *Streptomyces avermitilis*
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1389)
US-10-156-761-780

Query Match 28.8%; Score 15; DB 14; Length 1389;
Best Local Similarity 100.0%; Pred. No. 61;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 CGTCGCTGCCGCCCT 44
|||||
Db 19 CGTCGCTGCCGCCCT 5

RESULT 28

US-10-177-293-416/c
; Sequence 416, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James

```

; APPLICANT: Glat, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Ganavarpur Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puzstai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 416
; LENGTH: 1404
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-177-293-416

Query Match      28.8%; Score 15; DB 14; Length 1404;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      19 CTCATCGCTCTGCTC 33
Db      122 CTCATCGCTCTGCTC 108

RESULT 29
US-10-156-761-4461/c
; Sequence 4461, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4461

```

```

; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1479)
US-10-156-761-4461

Query Match      28.8%; Score 15; DB 14; Length 1479;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      27 TCTCGTGGTGGCCGG 41
Db      1258 TCTCGTGGTGGCCGG 1244

RESULT 30
US-10-128-714-7070/c
; Sequence 7070, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Erosnkin, Alexey M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7070
; LENGTH: 1554
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-7070

Query Match      28.8%; Score 15; DB 14; Length 1554;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      24 CGCTCTGCTGCTGCTC 38
Db      1384 CGCTCTGCTGCTGCTC 1370

RESULT 31
US-10-128-714-6070/c
; Sequence 6070, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Erosnkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: Methods of Use

```



```

# CURRENT APPLICATION NUMBER: US 60/285,999
# PRIOR FILING DATE: 2002-04-23
# CURRENT APPLICATION NUMBER: US 60/285,997
# PRIOR FILING DATE: 2001-04-23
# PRIOR APPLICATION NUMBER: US 60/287,066
# PRIOR FILING DATE: 2001-04-27
# PRIOR APPLICATION NUMBER: US 60/295,890
# PRIOR FILING DATE: 2001-06-05
# PRIOR APPLICATION NUMBER: US 60/303,899
# PRIOR FILING DATE: 2001-07-09
# PRIOR APPLICATION NUMBER: US 60/316,362
# PRIOR FILING DATE: 2001-08-31
# NUMBER OF SEQ ID NOS: 8603
# SOFTWARE: PatentIn version 3.1
# SEQ ID NO: 6070
# LENGTH: 1778
# TYPE: DNA
# ORGANISM: Aspergillus fumigatus
US-10-128-714-6070
```

```
RESULT 35
US-09-815-242-4138
; Sequence 4138, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4138
; LENGTH: 2454
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4138

Query Match      28.8%; Score 15; DB 9; Length 2454;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      30 CGTCGCTGCGGCGCT 44
Db      12 CGTCGCTGCGGCGCT 26

RESULT 36
US-09-374-046A-125
; Sequence 125, Application US/09374046A
; Publication No. US20030096951A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Werberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steindinger II, Robert J.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fecthel, Kim
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: GI 6075-83A
; CURRENT APPLICATION NUMBER: US/09/374,046A
```

```
; CURRENT FILING DATE: 1999-08-13
; NUMBER OF SEQ ID NOS: 240
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 2672
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-374-046A-125

Query Match      28.8%; Score 15; DB 11; Length 2672;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      17 CTCCTCATCGCTCCG 31
Db      153 CTCCTCATCGCTCCG 167

RESULT 37
US-09-954-456-518/C
; Sequence 518, Application US/09954456
; Patent No. US2002015057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Car
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 518
; LENGTH: 2790
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-518

Query Match      28.8%; Score 15; DB 10; Length 2790;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      15 CGCTTCATCGCTCT 29
Db      984 CGCTTCATCGCTCT 970

RESULT 38
US-09-967-768A-182/C
; Sequence 182, Application US/09967768A
; Patent No. US20020150877A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signat
```

```

; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 182
; LENGTH: 2790
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-967-768A-182

Query Match      28.8%; Score 15; DB 10; Length 2790;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      15  CGCTCTCATGCTCT 29
Db      984 CGCTCTCATGCTCT 970

RESULT 39
US-10-128-714-70/C
; Sequence 70, Application US/10128714
; Publication NO. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshekin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 3140
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
; US-10-128-714-70

Query Match      28.8%; Score 15; DB 14; Length 3140;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      24  CGCTCTCGTCTGCTG 38
Db      2608 CGCTCTCGTCTGCTG 2594

RESULT 40
US-09-822-846-197
; Sequence 197, Application US/09822846

```

```

; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fectel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakari
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 197
; LENGTH: 3323
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-822-846-197

Query Match      28.8%; Score 15; DB 11; Length 3323;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      17  CTCTCATGCTCTG 31
Db      153 CTCTCATGCTCTG 167

Search completed: November 13, 2003, 11:54:39
Job time : 246 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 09:41:06 ; Search time 198 Seconds

(without alignments)
648.785 Million cell updates/sec

Title: US-10-081-935-1

Perfect score: 52

Sequence: 1 atgcagctctctcagcctc.....cgctgcgcgcgcgcgcgcgcg 52

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gsa_hum:*
18: em_gsa_inv:*
19: em_gsa_pin:*
20: em_gsa_vrt:*
21: em_gsa_fun:*
22: em_gsa_mam:*
23: em_gsa_mus:*
24: em_gsa_pro:*
25: em_gsa_rtd:*
26: em_gsa_phg:*
27: em_gsa_vhl:*
28: gb_gsa1:*
29: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	395	10	BF251279 EST418539
2	52	100.0	401	10	BF252856 EST420119
3	52	100.0	531	10	BF251334 EST418594
4	52	100.0	557	10	BF252723 EST419986

5	52	100.0	737	10	BF251103	BF251103 EST418360
6	23	44.2	311	13	BQ499065	BQ499065 EST088290
7	19	36.5	133	12	BP112708	BP112708 BP112708
8	19	36.5	314	9	AUI63553	AUI63553 AUI63553
9	19	36.5	460	9	AUI66766	AUI66766 AUI66766
10	19	36.5	468	9	AUI031845	AUI031845 AUI031845
11	19	36.5	585	28	AQ510461	AQ510461 nbxb00958
12	19	36.5	758	14	CB684405	CB684405 OSJNE0130
13	19	36.5	767	14	CB939902	CB939902 IPCGX14
14	19	36.5	978	10	BE039817	BE039817 OC08E11
15	19	36.5	1712	10	BG163853	BG163853 602343204
16	18	34.6	368	28	AQ845447	AQ845447 LMAJFV1.1
17	18	34.6	546	12	B1568671	B1568671 RH40321.5
18	18	34.6	558	12	B1583892	B1583892 RH22614.5
19	18	34.6	610	12	B1626323	B1626323 RH67006.5
20	18	34.6	936	28	AZ208101	AZ208101 SF_0136.B
21	18	34.6	1012	29	CNS01736	AL107532 Drosoph11
22	18	34.6	1083	10	BF307868	BF307868 601890683
23	18	34.6	1101	29	CNS00ATX	AL056234 Drosoph11
24	18	34.6	1101	29	CNS014JM	AL104246 Drosoph11
25	17	32.7	343	12	BT123995	BT123995 1033P32P
26	17	32.7	359	14	CD262900	CD262900 pMAU19XK
27	17	32.7	374	13	BY076213	BY076213 BY076213
28	17	32.7	404	9	BF008574	BF008574 1613456.A
29	17	32.7	406	10	AM465276	AM465276 BP230018B
30	17	32.7	425	9	AM056821	AM056821 ST56P10.P
31	17	32.7	446	9	AM376507	AM376507 IL3-CT021
32	17	32.7	470	9	AM056821	AM056821 ST56P10.P
33	17	32.7	490	13	BQ170403	BQ170403 WHE1770.E
34	17	32.7	505	10	BE484921	BE484921 171741.BA
35	17	32.7	506	13	BO821431	BO821431 1030092C0
36	17	32.7	529	9	A1988643	A1988643 gd05H05.Y
37	17	32.7	542	9	AV668034	AV668034 AV668034
38	17	32.7	544	9	A1542489	A1542489 S008831.5
39	17	32.7	546	9	AV597173	AV597173 AV597173
40	17	32.7	556	9	AV603191	AV603191 AV603191
41	17	32.7	570	9	AV614001	AV614001 AV614001
42	17	32.7	598	9	AV662793	AV662793 AV662793
43	17	32.7	604	28	AQ946593	AQ946593 Sheared.D
44	17	32.7	616	28	BH335108	BH335108 CH230-99G
45	17	32.7	626	13	BU044311	BU044311 PP_LEA001
46	17	32.7	663	10	BE776572	BE776572 MY-17-F-1
47	17	32.7	663	13	BQ162104	BQ162104 WHE0496.E
48	17	32.7	668	14	CD258920	CD258920 p1MY0175F
49	17	32.7	677	29	BZ363829	BZ363829 1d91C10.G
50	17	32.7	694	14	CD491843	CD491843 WHE2222.G
51	17	32.7	707	14	CB419666	CB419666 592540.MA
52	17	32.7	758	29	BZ360194	BZ360194 1d91C10.b
53	17	32.7	760	14	CB653954	CB653954 OSJNEC05H
54	17	32.7	770	10	BF865131	BF865131 963057D04
55	17	32.7	771	29	CNS017FC	AL166449 Tetraodon
56	17	32.7	812	14	BT728779	BT728779 BY728779
57	17	32.7	822	29	CNS0421Q	AL1271547 Tetraodon
58	17	32.7	839	10	BG701989	BG701989 602683363
59	17	32.7	960	29	CNS01EMC	AL140725 Anopheles
60	17	32.7	978	29	CNS078HZ	AL134013 T3 end of
61	17	32.7	979	12	BM470883	BM470883 AGENCOURT
62	17	32.7	1013	29	CNS03K1A	AL156663 Tetraodon
63	17	32.7	1099	29	BZ579805	AL256663 Tetraodon
64	17	32.7	1201	13	BX448656	BX448656 BX448656
65	17	32.7	1269	10	BE963132	BE963132 601656936
66	17	32.7	1654	29	AC082054	AC082054 Pan.trog1
67	17	32.7	1947	10	BF303726	BF303726 602032601
68	16	30.8	113	9	AL825153	AL825153
69	16	30.8	172	9	AA665646	AA665646 ag55F03..8
70	16	30.8	188	29	DME545703	AJ4545703 Drosoph11
71	16	30.8	207	16	AG215251	AG215251 Drosoph11
72	16	30.8	262	29	AJ497920	AJ497920
73	16	30.8	263	28	AQ911925	AQ911925 LMAJFV1.1
74	16	30.8	284	28	BH904772	BH904772 SALK_1050
75	16	30.8	312	29	CNS01K1L	AL1448490 Anopheles
76	16	30.8	321	13	BU644056	BU644056 mgmK018XO
77	16	30.8	334	12	BMS77157	BMS77157 170006871


```
C 954 14 26.9 237 9 A1988678 A1988678 sdo6e07.y
C 955 14 26.9 238 10 BE506373 BE506373 db82c08.y
C 956 14 26.9 239 29 CC001754 CC001754 PUDHT907D
C 957 14 26.9 241 29 B2628644 B2628644 ih6id10.b
C 958 14 26.9 243 9 A1621491 A1621491 486092A08
C 959 14 26.9 243 13 BQ163172 BQ163172 952045G01
C 960 14 26.9 243 13 B170223 B170223 B170223 B170223
C 961 14 26.9 245 28 BH406412 BH406412 RPCI-23-4
C 962 14 26.9 247 10 BG658056 BG658056 T9ESTZY44
C 963 14 26.9 247 13 BU546528 BU546528 CM880008B
C 964 14 26.9 248 9 A1284046 A1284046 4AAB-AAT-
C 965 14 26.9 248 9 AV431283 AV431283 AV431283
C 966 14 26.9 248 13 BY073743 BY073743 BY073743
C 967 14 26.9 249 28 BH832799 BH832799 BACPP4-E1
C 968 14 26.9 249 28 BH833764 BH833764 BACPP5-H2
C 969 14 26.9 250 10 BF082559 BF082559 OVI-BT063
C 970 14 26.9 251 13 BU660029 BU660029 C155A09.z
C 971 14 26.9 253 13 BU585451 BU585451 A1-am-05G
C 972 14 26.9 254 14 CA890022 CA890022 B0156E11-
C 973 14 26.9 256 14 CB126527 CB126527 K-EST0175
C 974 14 26.9 259 10 BE497308 BE497308 WHE0766.F
C 975 14 26.9 259 12 B1543121 B1543121 949074D01
C 976 14 26.9 262 28 BH617680 BH617680 SALK_0374
C 977 14 26.9 266 10 BE530068 BE530068 M76B04STM
C 978 14 26.9 267 9 AM706963 AM706963 sk08h01.y
C 979 14 26.9 267 10 BG383097 BG383097 J01028.MA
C 980 14 26.9 269 12 BJ300206 BJ300206 BJ300206
C 981 14 26.9 270 9 AA597456 AA597456 29727.Lam
C 982 14 26.9 272 9 AM000380 AM000380 614014D10
C 983 14 26.9 276 10 BB242079 BB242079 BB242079
C 984 14 26.9 277 10 BB862181 BB862181 BB862181
C 985 14 26.9 277 13 BY680653 BY680653 BY680653
C 986 14 26.9 279 9 AV242658 AV242658 AV242658
C 987 14 26.9 280 12 BM745914 BM745914 K-EST0019
C 988 14 26.9 281 9 AA093779 AA093779 c10921.se
C 989 14 26.9 282 14 CD527261 CD527261 3529.1.11
C 990 14 26.9 282 13 BY191847 BY191847 BY191847
C 991 14 26.9 282 12 AZ351662 AZ351662 IM0089M18
C 992 14 26.9 284 12 BJ184962 BJ184962 BJ184962
C 993 14 26.9 285 9 AJ285427 AJ285427 4A3B-AAJ-
C 994 14 26.9 285 10 AM862388 AM862388 RC0-CT038
C 995 14 26.9 285 13 BY237153 BY237153 BY237153
C 996 14 26.9 285 28 BH870227 BH870227 hm58f06.b
C 997 14 26.9 286 14 CB346362 CB346362 CB346362
C 998 14 26.9 291 13 BQ345037 BQ345037 CM2-N7016
C 999 14 26.9 291 14 CD477528 CD477528 eca01-27m
C 1000 14 26.9 292 12 BG933341 BG933341 WS1_3_P02
```

ALIGNMENTS

```
RESULT 1 395 bp mRNA linear EST 15-NOV-2001
LOCUS BF251279
DEFINITION BF251279 Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAAfl1 5' sequence, mRNA sequence.
ACCESSION BF251279.1 GI:16931422
VERSION BF251279.1
KEYWORDS EST.
SOURCE Coccidioides immitis
ORGANISM Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Omygenales; mitosporic Omygenales; Coccidioides.
REFERENCE 1 (bases 1 to 395)
AUTHORS Gardner M.J. and Kirkland, T.
TITLE Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL unpublished
COMMENT Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
```

```
Fax: 301 838 0208
Email: gardner@igir.org.
FEATURES
Location/Qualifiers
source 1..395
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAAfl1"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_11b="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 78 a 133 c 78 g 106 t
ORIGIN
Query Match 100.0%; Score 52; DB 10; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.1e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGTTCTCTACGCTCTCATGCTCTGCTGCTGCGGCTCGGCTCGGCTG 52
142 ATGCAGTTCTCTACGCTCTCATGCTCTGCTGCTGCTGCGGCTCGGCTG 193
Db
```

```
RESULT 2 401 bp mRNA linear EST 15-NOV-2001
LOCUS BF252856
DEFINITION BF252856 Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAAB70 5' sequence, mRNA sequence.
ACCESSION BF252856.1 GI:16932999
VERSION BF252856
KEYWORDS EST.
SOURCE Coccidioides immitis
ORGANISM Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Omygenales; mitosporic Omygenales; Coccidioides.
REFERENCE 1 (bases 1 to 401)
AUTHORS Gardner M.J. and Kirkland, T.
TITLE Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL unpublished
COMMENT Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@igir.org.
```

FEATURES

```
source 1..401
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAAB70"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_11b="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 77 a 149 c 89 g 86 t
ORIGIN
Query Match 100.0%; Score 52; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 3.1e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGTTCTCTACGCTCTCATGCTCTGCTGCTGCGGCTCGGCTCGGCTG 52
Db 59 ATGCAGTTCTCTACGCTCTCATGCTCTGCTGCTGCTGCGGCTCGGCTG 110
RESULT 3 BF251334
```

LOCUS BF251334 531 bp mRNA linear EST 15-NOV-2001
DEFINITION EST148594 Coccidioides immitis spherule cDNA library Coccidioides
ACCESSION BF251334
VERSION BF251334.1 GI:16931477
KEYWORDS EST.
SOURCE Coccidioides immitis
ORGANISM Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE 1 (bases 1 to 531)
AUTHORS Gardner, M.J. and Kirkland, T.
TITLE Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL Unpublished
COMMENT Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org.
Location/Qualifiers
1..531
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAAF68"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 107 a 190 c 115 g 119 t
ORIGIN
Query Match 100.0%; Score 52; DB 10; Length 531;
Best Local Similarity 100.0%; Pred. No. 3.3e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGTTCTTCACGCTTCATCGCTCTGCTGCGTCCGCGCTCCGCACTG 52
Db 137 ATGCAGTTCTTCACGCTTCATCGCTCTGCTGCGTCCGCGCTCCGCACTG 188

RESULT 4
LOCUS BF252723 557 bp mRNA linear EST 15-NOV-2001
DEFINITION EST149986 Coccidioides immitis spherule cDNA library Coccidioides
ACCESSION BF252723
VERSION BF252723.1 GI:16932866
KEYWORDS EST.
SOURCE Coccidioides immitis
ORGANISM Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE 1 (bases 1 to 557)
AUTHORS Gardner, M.J. and Kirkland, T.
TITLE Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL Unpublished
COMMENT Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org.
Location/Qualifiers
1..557
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAA21"

FEATURES
source

/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 110 a 197 c 117 g 133 t
ORIGIN
Query Match 100.0%; Score 52; DB 10; Length 557;
Best Local Similarity 100.0%; Pred. No. 3.3e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGTTCTTCACGCTTCATCGCTCTGCTGCGTCCGCGCTCCGCACTG 52
Db 161 ATGCAGTTCTTCACGCTTCATCGCTCTGCTGCGTCCGCGCTCCGCACTG 212

RESULT 5
LOCUS BF251103 737 bp mRNA linear EST 15-NOV-2001
DEFINITION EST148360 Coccidioides immitis spherule cDNA library Coccidioides
ACCESSION BF251103
VERSION BF251103.1 GI:16931246
KEYWORDS EST.
SOURCE Coccidioides immitis
ORGANISM Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE 1 (bases 1 to 737)
AUTHORS Gardner, M.J. and Kirkland, T.
TITLE Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL Unpublished
COMMENT Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org.
Location/Qualifiers
1..737
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAAC67"
/dev_stage="spherule"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 136 a 271 c 168 g 162 t
ORIGIN
Query Match 100.0%; Score 52; DB 10; Length 737;
Best Local Similarity 100.0%; Pred. No. 3.5e-17;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGTTCTTCACGCTTCATCGCTCTGCTGCGTCCGCGCTCCGCACTG 52
Db 99 ATGCAGTTCTTCACGCTTCATCGCTCTGCTGCGTCCGCGCTCCGCACTG 150

RESULT 6
LOCUS BQ499065 311 bp mRNA linear EST 31-OCT-2002
DEFINITION EST08290 Pb0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.
ACCESSION BQ499065
VERSION BQ499065.1 GI:24452039
KEYWORDS EST.
SOURCE Paracoccidioides brasiliensis
ORGANISM Paracoccidioides brasiliensis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

REFERENCE 1 (bases 1 to 311)
 Oxyenales; mitosporic Oxyenales; Paracoccidioides.
 GOLDMAN, G.H., MARQUES, E.R., RIBEIRO, D.C.D., BERNARDES, L.A., PUCCIA, R., TRAVASSOS, L.R., NOBREGA, F.G., NOBREGA, M.P., SAVOLDI-BARBOSA, M., SEMIGHINI, C.P. and GOLDMAN, M.H.
 The Paracoccidioides brasiliensis EST genome project
 Eukaryotic Cell, (2002) in press
 CONTACT: Gustavo Henrique Goldman
 Laboratório de Molecular Biology
 Universidade de São Paulo - USP - FCFRP
 Av. do Café S/N, CEP: 14040-903, Ribeirão Preto - SP, Brazil
 Email: gsgoldman@usp.br.
 Location/Qualifiers

FEATURES
 source 1..311
 /organism="Paracoccidioides brasiliensis"
 /mol_type="mRNA"
 /db_xref="taxon:121759"
 /clone_lib="PB0001"

BASE COUNT 73 a 88 c 59 g 91 t

ORIGIN

Query Match 44.2%; Score 23; DB 13; Length 311;
 Best Local Similarity 100.0%; Pred. No. 0.054;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGCAGTTCTCTCAGCTCTCAT 23
 |||||
 91 ATGCAGTTCTCTCAGCTCTCAT 113

RESULT 7
 BP112708/c 133 bp mRNA linear EST 11-FEB-2003
 DEFINITION BP112708 ORCS bovine utero-placenta cDNA Bos taurus CDNA clone
 ORCS13477 5', mRNA sequence.
 BP112708
 BP112708.1 GI:28314998
 EST.
 Bos taurus (cow)
 Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 133)
 Ishiwata, H., Katsuna, S., Kizaki, K., Patel, O.V., Nakano, H.,
 Takahashi, T., Imai, K., Hirasawa, A., Shiojima, S., Ikawa, H., Suzuki,
 Y., Tsujimoto, G., Izaike, Y., Todoroki, J. and Hashizume, K.
 Characterization of gene expression profiles in early bovine
 pregnancy using a custom cDNA microarray
 Mol. Reprod. Dev. 65 (1), 9-18 (2003)
 22544902
 12658628

TITLE
 JOURNAL
 MEDLINE
 PUBMED

COMMENT
 Contact: Gozoh Tsujimoto
 Department of Molecular, Cell Pharmacology
 National Research Institute for Child Health and Development
 3-35-31 Taishido, Setagaya, Tokyo 154-8567, Japan
 Tel: 81-3-3149-2476
 Fax: 81-3-3149-1252
 Email: gtsujimoto@nch.go.jp
 This work was performed to collaborate with Developmental Biology
 Department, National Institute of Agricultural Sciences. Address:
 2 Ikenodai, Tsukuba, Ibaraki, 305-8602 Japan. Phone & Fax:
 81-29-838-8633 e-mail: kazuha@affrc.go.jp
 This work was funded by Organized Research Combination System (ORCS
) Project of Ministry of Education, Culture, Sports, Science and
 Technology.
 Location/Qualifiers

FEATURES
 source 1..133
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /clone="ORCS13477"
 /tissue_type="mixture of uterus and placenta"

BASE COUNT 34 a 33 c 29 g 37 t

ORIGIN

Query Match 36.5%; Score 19; DB 12; Length 133;
 Best Local Similarity 100.0%; Pred. No. 5.8;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 CAGTTCCTCAGCTCTCA 22
 |||||
 67 CAGTTCCTCAGCTCTCA 49

RESULT 8
 AU163553 314 bp mRNA linear EST 03-APR-2002
 LOCUS AU163553 Rice panicle at flowering stage Oryza sativa (japonica
 DEFINITION AU163553 Rice panicle at flowering stage Oryza sativa (japonica
 cultivar-group) cDNA clone E1586, mRNA sequence.
 AU163553
 AU163553.1 GI:11170958
 EST.
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Euhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 314)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from panicle at flowering stage (2000)
 Unpublished
 Contact: Takuji Sasaki
 National Institute of Agricultural Resources
 Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
 305-8602, Japan
 Tel: 81-298-38-7441
 Fax: 81-298-38-7468
 Email: tsasaki@affrc.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
 PROJECT "RGP".
 Location/Qualifiers

FEATURES
 source 1..314
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="E1586"
 /dev_stage="flowering stage"
 /clone_lib="Rice panicle at flowering stage"
 /note="Organ: panicle; Rice cDNA from panicle at flowering
 stage"

BASE COUNT 27 a 120 c 107 g 57 t 3 others

ORIGIN

Query Match 36.5%; Score 19; DB 9; Length 314;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 CGCTCTCGTCCGCTCCGCGC 42
 |||||
 187 CGCTCTCGTCCGCTCCGCGC 205

RESULT 9
 AU166766/c 460 bp mRNA linear EST 03-APR-2002
 LOCUS AU166766 Rice callus (2001) Oryza sativa (japonica cultivar-group)
 DEFINITION AU166766 Rice callus (2001) Oryza sativa (japonica cultivar-group)
 cDNA clone C50247, mRNA sequence.
 AU166766
 AU166766.1 GI:12405165
 EST.
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 460)
Sasaki, T. and Yamamoto, K.
Rice cDNA from callus (2001)
Unpublished
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP'.

FEATURES
source
1..460
Location/Qualifiers

1.460
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="C50247"
/cissue_type="callus"
/clone_1ib="Rice callus (2001)"
110 a 123 c 149 g 76 t 2 others

BASE COUNT
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 19; DB 9; Length 460;
Pred. No. 7.5; Mismatches 0; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 CGCTCTGCTGCTGCGCGC 42
|||||
88 CGCTCTGCTGCTGCGCGC 70

RESULT 10
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AU031845 468 bp mRNA linear EST 01-APR-2002
AU031845 Rice root Oryza sativa (japonica cultivar-group) cDNA
clone R2118_62, mRNA sequence.
AU031845
AU031845.1 GI:3767818
EST.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 468)
Minobe, Y. and Sasaki, T.
Rice cDNA from root
Unpublished
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP'
POLYA=No.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1.468
Location/Qualifiers

1.468
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="R2118_62"
/clone_1ib="Rice root"
/note="Prepared from seedling root."

BASE COUNT
ORIGIN
130 a 99 c 118 g 121 t

Query Match
Best Local Similarity 100.0%; Score 19; DB 9; Length 468;
Pred. No. 7.6; Mismatches 0; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 14 ACGCTCTCATCGCTCTGCT 32
|||||
Db 101 ACGCTCTCATCGCTCTGCT 83

RESULT 11
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

A0510461 585 bp DNA linear GSS 04-MAY-1999
nbxb0095B03r CUGI Rice BAC library Oryza sativa (japonica
cultivar-group) genomic clone nbxb0095B03r, genomic survey
sequence.
A0510461
A0510461.1 GI:4733065
GSS.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 585)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GGAACACGCTATGACCATG
Clase: BAC ends
High quality sequence stop: 467.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: GGAACACGCTATGACCATG
Clase: BAC ends
High quality sequence stop: 467.

FEATURES
source
1..585
Location/Qualifiers

1.585
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/strain="japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nbxb0095B03r"
/cissue_type="leaf"
/lab_host="E. coli DH10B"
/clone_1ib="CUGI Rice BAC library"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9%. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."

BASE COUNT
ORIGIN
110 a 203 c 150 g 122 t

Query Match
Best Local Similarity 100.0%; Score 19; DB 28; Length 585;
Pred. No. 7.9; Mismatches 0; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 CTATCGCTCTCGTCGCTG 37
 |||||
 DB 160 CTCATCGCTCTCGTCGCTG 178

RESULT 12
 CB684405 758 bp mRNA linear EST 09-APR-2003
 LOCUS OSJNEF13003.3 OSJNEF Oryza sativa (japonica cultivar-group) cDNA
 DEFINITION
 CB684405
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 EST.
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Erihartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 758)
 Jantanasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
 Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aaa cga cgg cca gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 13 row: 0 column: 03
 Seq primer: gga aac agc tat gac cat g.
 Location/Qualifiers
 1..758
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEF13003"
 /issue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSJNEF"
 /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
 XhoI; Uninfected Control"

BASE COUNT 137 a 246 c 198 g 177 t

ORIGIN

Query Match 36.5%; Score 19; DB 14; Length 758;
 Best Local Similarity 100.0%; Pred. No. 8.4;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 CGCTCTCGTCGCTCGCGC 42
 |||||
 DB 406 CGCTCTCGTCGCTCGCGC 424

RESULT 13
 CB939902 767 bp mRNA linear EST 29-APR-2003
 LOCUS IPCGJx14.9 G08.23 IPCGJx14 Ictalurus punctatus cDNA clone
 DEFINITION
 IPCGJx14.9 G08.5', mRNA sequence.
 CB939902
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 EST.
 Ictalurus punctatus (channel catfish)
 Ictalurus punctatus

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Siluriformes;
 Ictaluridae; Ictalurus.
 1 (bases 1 to 767)
 Bengten,E., Khayat,M., Middleton,D., Waldbieser,G., Askovic,S.,
 Jensen,K.T., Marr,G., Miller,N., Cizm,L.W. and Wilson,M.
 Identification of expressed genes in mixed leukocyte culture and in
 a macrophage cell line in channel catfish, Ictalurus punctatus
 Unpublished
 Contact: Waldbieser GC
 Catfish Genetics Research Unit
 USDA-Agricultural Research Service
 141 Experiment Station Road, Stoneville, MS 38776, USA
 Tel: 662 686 3593
 Fax: 662 686 3567
 Email: gwaldbieser@ars.usda.gov
 Single pass sequencing. Bases called with Phred v0.000925.c. Low
 quality bases and vector trimmed with Lucy v1.16.
 Plate: 9 row: G column: 8
 Seq primer: T7.
 Location/Qualifiers
 1..767
 /organism="Ictalurus punctatus"
 /mol_type="mRNA"
 /strain="commercial"
 /db_xref="taxon:7998"
 /clone="IPCGJx14.9 G08"
 /cell_type="macrophage"
 /cell_line="42TA"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="IPCGJx14"
 /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
 Primary library"

BASE COUNT 250 a 169 c 271 g 77 t

ORIGIN

Query Match 36.5%; Score 19; DB 14; Length 767;
 Best Local Similarity 100.0%; Pred. No. 8.4;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 GCTCTCATCGCTCTCGTCG 34
 |||||
 DB 506 GCTCTCATCGCTCTCGTCG 488

RESULT 14
 BE039817 978 bp mRNA linear EST 07-JUN-2000
 LOCUS OC08E11 OC Oryza sativa cDNA 5', mRNA sequence.
 DEFINITION
 BE039817
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 EST.
 Oryza sativa
 Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Erihartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 978)
 Bohmert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferreira
 ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
 Scara,G., Wheeler,M. and Zepeda,G.R.
 Functional Genomics of Plant Stress Tolerance
 Unpublished
 Contact: Michalowski,C.B.
 University of Arizona
 Bio Sciences West room 513, Tucson, AZ 85721, USA
 Tel: 520-621-7982
 Fax: 520-621-1697
 Email: cbm@u.arizona.edu
 Best blastx match: 'emb|CA16536.1| (A1021633) predicted protein
 [Arabidopsis thaliana] 51 1e-05'. An open reading frame exists.
 Location/Qualifiers

FEATURES

source

1. .978
 /organism="Oryza sativa"
 /mol_type="mRNA"
 /strain="pokkali"
 /db_xref="taxon:4530"
 /tissue_type="roots"
 /dev_stage="1 week"
 /clone_1lb="OC"
 /note="no stress"

BASE COUNT 232 a 188 c 313 g 244 t 1 others
 ORIGIN

Query Match 36.5%; Score 19; DB 10; Length 978;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 ACCTCTCATCGCTCTCGT 32
 |||||
 Db 168 ACCTCTCATCGCTCTCGT 150

RESULT 15

LOCUS BGI63853 1712 bp mRNA linear EST 06-FEB-2001
 DEFINITION 602343204F1 NIH_MGC_89 Homo sapiens CDNA clone IMAGE:4453297 5',
 mRNA sequence.
 ACCESSION BGI63853
 VERSION BGI63853.1 GI:12670556
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1712)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM10243 row: h column: 02
 High quality sequence stop: 5.
 Location/Qualifiers

FEATURES

1. .1712
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_1lb="IMAGE:4453297"
 /tissue_type="hypertrophoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="NIH MGC 89"
 /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.3 Kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

BASE COUNT 691 a 415 c 350 g 256 t
 ORIGIN

Query Match 36.5%; Score 19; DB 10; Length 1712;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CGCTCTCATCGCTCTCGT 33
 |||||
 Db 1632 CGCTCTCATCGCTCTCGT 1614

RESULT 16

LOCUS A0845447/c 368 bp DNA linear GSS 25-MAY-2001
 DEFINITION LMAJFV1_lm25f07.y1 Leishmania major FV1 random genomic library
 Leishmania major genomic clone LMAJFV1_lm25f07 5', genomic survey
 sequence.
 ACCESSION A0845447 GI:6050095
 VERSION A0845447
 KEYWORDS GSS.
 SOURCE Leishmania major
 ORGANISM Leishmania major
 Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 368)
 Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L.,
 Kissinger, J.C., Roos, D.S., and Beverley, S.M.
 A survey of the Leishmania major Friedlin strain V1 genome by
 shotgun sequencing: a resource for DNA microarrays and expression
 profiling

Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
 MEDLINE 21192569
 PUBMED 11295190
 Other GSSs: lm25f07.x1
 Contact: Akopyants, NS / Beverley, SM
 Washu Leishmania Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: eatw@wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 If using this information please cite:
 N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
 Friedlin strain V1 genome by shotgun sequencing' and the Washington
 University Genome Sequencing Center for information on obtaining
 clone material please contact: Natalia S. Akopyants Ph.D.
 (natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
 (beverley@borcim.wustl.edu)
 Seq primer: -40RP from Gibco
 Class: shotgun
 High quality sequence stop: 358.
 Location/Qualifiers

FEATURES

1. .368
 Location/Qualifiers
 /organism="Leishmania major"
 /mol_type="genomic DNA"
 /strain="Friedlin strain V1"
 /db_xref="taxon:5664"
 /clone="LMAJFV1_lm25f07"
 /lab_host="TOP10 (Invitrogen)"
 /clone_1lb="Leishmania major FV1 random genomic library"
 /note="Vector: pZero-2 (Invitrogen); Site 1: EcoRV;
 Genomic DNA was isolated from stationary phase cells. For
 this library, DNA was sheared to give a tight size
 distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
 polymerase, dephosphorylated with Shrimp Alkaline
 Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 90 a 94 c 123 g 61 t
 ORIGIN

Query Match 34.6%; Score 18; DB 28; Length 368;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 TCATCGCTCTCGCTGCTG 37
 |||||
 Db 135 TCATCGCTCTCGCTGCTG 118

RESULT 17

LOCUS B1568671/c 546 bp mRNA linear EST 06-SEP-2001

DEFINITION RH40331.5prime RH Drosophila melanogaster normalized Head pflc-1
ACCESSION Drosophila melanogaster cDNA clone RH40331 5, mRNA sequence.
VERSION B1568671
KEYWORDS B1568671.1 GI:15460093
SOURCE EST
ORGANISM Drosophila melanogaster (fruit fly)
REFERENCE Drosophila melanogaster
AUTHORS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyroidae; Drosophilidae; Drosophila.
1 (bases 1 to 546)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
J., Champagne, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Miera, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phonenuavong, S., Wan, K., Yu, C., Lewis, S.E., Celnikier, S. and Rubin
G.M.
BDGP/HMI RH Drosophila EST Project
JOURNAL Unpublished
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AEO03466: arm:2R [19937589,20220687]
estimated-cyto:60E8-60F4: 08/18/2001
Plate: RH.403 row: C column: 7
High quality sequence stop: 508.
Location/Qualifiers
1..546
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RH40331"
/sex="male and female"
/dev_stage="Adult"
/lab_host="DHS-alpha Tona"
/clone_lib="RH Drosophila melanogaster normalized Head
pflc-1"
/note="Organ: head; Vector: pflc1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 100 a 148 c 176 g 120 t 2 others
ORIGIN
Query Match 34.6%; Score 18; DB 12; Length 546;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 CATGCTCTCGTCGCTGC 38
|||||
231 CATGCTCTCGTCGCTGC 214
|||||
RESULT 18
B1568671/c 558 bp mRNA linear EST 06-SEP-2001
LOCUS RH22614.5prime RH Drosophila melanogaster normalized Head pflc-1
DEFINITION Drosophila melanogaster cDNA clone RH22614 5, mRNA sequence.
ACCESSION B1568671
VERSION B1568671.1 GI:15475314
KEYWORDS EST
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyroidae; Drosophilidae; Drosophila.
1 (bases 1 to 558)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
J., Champagne, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Miera, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phonenuavong, S., Wan, K., Yu, C., Lewis, S.E., Celnikier, S. and Rubin
G.M.
BDGP/HMI RH Drosophila EST Project
JOURNAL Unpublished
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AEO03786: arm:2R [607836,924593]
estimated-cyto:41F5-41F9: 08/23/2001
Plate: RH.670 row: A column: 6

TITLE Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phonenuavong, S., Wan, K., Yu, C., Lewis, S.E., Celnikier, S. and Rubin
G.M.
JOURNAL BDGP/HMI RH Drosophila EST Project
COMMENT Unpublished
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AEO03786: arm:2R [607836,924593]
estimated-cyto:41F5-41F9: 08/17/2001
Plate: RH.226 row: B column: 2
High quality sequence stop: 501.
Location/Qualifiers
1..558
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RH22614"
/sex="male and female"
/dev_stage="Adult"
/lab_host="DHS-alpha Tona"
/clone_lib="RH Drosophila melanogaster normalized Head
pflc-1"
/note="Organ: head; Vector: pflc1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
BASE COUNT 101 a 152 c 180 g 124 t 1 others
ORIGIN
Query Match 34.6%; Score 18; DB 12; Length 558;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 CATGCTCTCGTCGCTGC 38
|||||
231 CATGCTCTCGTCGCTGC 214
|||||
RESULT 19
B1626323/c 610 bp mRNA linear EST 07-SEP-2001
LOCUS RH67006.5prime RH Drosophila melanogaster normalized Head pflc-1
DEFINITION Drosophila melanogaster cDNA clone RH67006 5, mRNA sequence.
ACCESSION B1626323
VERSION B1626323.1 GI:15521848
KEYWORDS EST
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyroidae; Drosophilidae; Drosophila.
1 (bases 1 to 610)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
J., Champagne, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Miera, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phonenuavong, S., Wan, K., Yu, C., Lewis, S.E., Celnikier, S. and Rubin
G.M.
BDGP/HMI RH Drosophila EST Project
JOURNAL Unpublished
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AEO03786: arm:2R [607836,924593]
estimated-cyto:41F5-41F9: 08/23/2001
Plate: RH.670 row: A column: 6

```

FEATURES
  source
    High quality sequence stop: 525.
    Location/Qualifiers
      1..610
        /organism="Drosophila melanogaster"
        /mol_type="mRNA"
        /db_xref="taxon:7227"
        /clone="RH67006"
        /sex="male and female"
        /dev_stage="Adult"
        /lab_host="DH5-alpha Tona"
        /clone_11b="RH Drosophila melanogaster normalized Head
        p1c-1"
        /note="Organ: head; Vector: pFlc1; Site: 1; XhoI; Site 2:
        BamHI; Library was kindly generated by Piero Carninci at
        the RIKEN. The library was normalized and excised using
        Cre recombinase. Plasmid cDNA library."
BASE COUNT
  115 a 163 c 198 g 133 t 1 others
ORIGIN
  Query Match
  Best Local Similarity 100.0%; Pred. No. 27;
  Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
  21 CATCGCTCTCGTCGCTGC 38
  |||||
  231 CATCGCTCTCGTCGCTGC 214

RESULT 20
LOCUS
  AZ208101 936 bp DNA linear GSS 31-AUG-2000
DEFINITION
  SP 0136_B1 F01 T7A Strongylocentrotus purpuratus, purple sea urchin
  'sperm genomic BAC library Strongylocentrotus purpuratus genomic
  clone Plate=136 Col=1 Row=L, genomic survey sequence.
ACCESSION
  AZ208101 GI:8420375
VERSION
  GSS.
KEYWORDS
  Strongylocentrotus purpuratus
SOURCE
  Strongylocentrotus purpuratus
  Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
  Echinoidae; Euechinoidae; Echinacea; Echinoidea;
  Strongylocentrotidae; Strongylocentrotus.
  1 (bases 1 to 936)
REFERENCE
  Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
  Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,
  G.A., Eitensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
  Hood,L.
  A sea urchin genome project: Sequence scan, virtual map, and
  additional resources
  Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
JOURNAL
  MEDLINE
  PUBMED
  10920195
COMMENT
  Contact: Cameron, RA, Davidson, EH, Hood, L
  Division of Biology 156-29
  California Institute of Technology
  Pasadena California 91125, USA
  Tel: (626) 395-8421
  Fax: (626) 793-3047
  Email: acameron@caltech.edu
  Plate: 136 row: L column: 1
  Seq primer: T7
  Class: BAC ends
  High quality sequence stop: 936.
  Location/Qualifiers
    1..936
      /organism="Strongylocentrotus purpuratus"
      /mol_type="genomic DNA"
      /db_xref="taxon:7668"
      /clone="Plate=136 Col=1 Row=L"
      /clone_11b="Strongylocentrotus purpuratus, purple sea
      urchin, sperm genomic BAC library"
      /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli
      DH10B"

```

```

BASE COUNT
  227 a 231 c 163 g 315 t
ORIGIN
  Query Match
  Best Local Similarity 100.0%; Pred. No. 29;
  Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
  18 TCTCATCGCTCTCGTCGC 35
  |||||
  Db 607 TCTCATCGCTCTCGTCGC 624

RESULT 21
LOCUS
  CNS01736 1012 bp DNA linear GSS 26-JUL-1999
DEFINITION
  Drosophila melanogaster genome survey sequence T7 end of BAC
  BACN17C24 of DrosBAC library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL107532
  AL107532.1 GI:5627836
VERSION
  GSS.
KEYWORDS
  Drosophila melanogaster (fruit fly)
SOURCE
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 1012)
REFERENCE
  Genoscope.
  Direct Submission
  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
  - Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaud at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBelobAC11.
  Location/Qualifiers
    1..1012
      /organism="Drosophila melanogaster"
      /mol_type="genomic DNA"
      /db_xref="taxon:7227"
      /clone="BACN17C24"
      /clone_11b="DrosBAC"
      /plasmid="pBelobAC11"
      /note="end : T7"
BASE COUNT
  192 a 287 c 249 g 233 t 51 others
ORIGIN
  Query Match
  Best Local Similarity 100.0%; Pred. No. 30;
  Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
  21 CATCGCTCTCGTCGCTGC 38
  |||||
  Db 953 CATCGCTCTCGTCGCTGC 970

RESULT 22
LOCUS
  BF307868 1083 bp mRNA linear EST 21-NOV-2000
DEFINITION
  601890683F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4131836 5',
  mRNA sequence.
ACCESSION
  BF307868
  BF307868.1 GI:11255033
VERSION
  EST.
KEYWORDS
  Homo sapiens (human)
SOURCE
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

REFERENCE 1 (bases 1 to 1083)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
 Plate: LNCM1030 row: e column: 21
 High quality sequence stop: 466.
 Location/Qualifiers

FEATURES

source

1. 1083
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4131836"
 /tissue_type="rhabdomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_11b="NIH MGC 17"
 /note="Organ: muscle; Vector: pONT7; Site 1: EcoRI; Site 2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 218 a 363 c 318 g 184 t
 ORIGIN

Query Match 34.6%; Score 18; DB 10; Length 1083;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 TGCAGTTCTCTCAGCGTC 19
 |||
 414 TGCAGTTCTCTCAGCGTC 431

RESULT 23
 CNS00ATX 1101 bp DNA linear GSS 04-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
 DEFINITION BACR22L05 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL056234.1 GI:4936801
 VERSION AL056234.1
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. 1101
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR22L05"
 /clone_11b="RPCI-98"
 /note="end : T7"
 BASE COUNT 254 a 295 c 253 g 226 t 73 others
 ORIGIN

Query Match 34.6%; Score 18; DB 29; Length 1101;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 21 CATCGCTCTGTCGTCGTC 38
 |||
 Db 608 CATCGCTCTGTCGTCGTC 625

RESULT 24
 CNS014JW 1101 bp DNA linear GSS 26-JUL-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
 DEFINITION BACN12F17 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL104246 GI:5615857
 VERSION AL104246
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES

source

1. 1101
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACN12F17"
 /clone_11b="DrosBAC"
 /plasmid="pBelobAC11"
 /note="end : T7"
 BASE COUNT 236 a 261 c 315 g 255 t 34 others
 ORIGIN

Query Match 34.6%; Score 18; DB 29; Length 1101;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 21 CATCGCTCTGTCGTCGTC 38
 |||
 Db 175 CATCGCTCTGTCGTCGTC 158

RESULT 25
LOCUS B1123995 343 bp mRNA linear EST 31-DEC-2001
DEFINITION 1033P32P Populus leaf cDNA library Populus tremula x Populus tremuloides cDNA, mRNA sequence.
ACCESSION B1123995
VERSION B1123995.1 GI:18007970
KEYWORDS EST.
SOURCE Populus tremula x Populus tremuloides
ORGANISM Populus tremula x Populus tremuloides
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
1 (bases 1 to 343)
Hertzberg, M., Aspeborg, H., Erlandsson, R., Bjorkbacka, H., Hilonen, T., Karlsson, J., Teeri, T., Gustafsson, P., Bahlert, R., Jansson, S., Nilsson, O., Sundberg, B., Nilsson, P., Uhlen, M., Sandberg, G. and Lundberg, J.
Gene expression in Populus
Unpublished
Contact: Erlandsson R
Department of Biotechnology
Royal Institute of Technology
Teknikringen 30, Stockholm S-10044, Sweden
Tel: 46 8 790 8287
Fax: 46 8 245452
Email: rikert@biochem.kth.se.

FEATURES
source
1..343
/organism="Populus tremula x Populus tremuloides"
/mol_type="mRNA"
/db_xref="taxon:47664"
/clone_lib="Populus leaf cDNA library"
/note="Organ: leaf"

BASE COUNT 75 a 89 c 58 g 121 t
ORIGIN

Query Match 32.7% Score 17; DB 12; Length 343;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 16 GCTCTCATCGCTCTCGT 32
|||||
6 GCTCTCATCGCTCTCGT 22

RESULT 26
LOCUS CD262900/c 359 bp mRNA linear EST 23-MAY-2003
DEFINITION p6MA019XK10F.192458 sMA: Phytophthora sojae grown in synthetic medium Phytophthora sojae cDNA clone sMA019K10.5, mRNA sequence.
ACCESSION CD262900
VERSION CD262900.1 GI:31050727
KEYWORDS EST.
SOURCE Phytophthora sojae
ORGANISM Phytophthora sojae
REFERENCE Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora
1 (bases 1 to 359)
Tyler, B.M., Judelson, H.S., Gijzen, M., Dean, R.A. and Waugh, M.E.
USDA-IFAFS: Expression of Phytophthora sojae genes during infection and propagation
Unpublished
Contact: Tyler B
Tyler lab
VBI
1890 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmtylet@vt.edu
PCR Primers
FORWARD: BK reverse
Plate: 019 row: K column: 10

Seq primer: BK reverse
High quality sequence scop: 359.
Location/Qualifiers
1..359
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"
/clone="sMA019K10"
/issue_type="mycelium"
/cell_line="P6497"
/dev_stage="mycelium"
/clone_lib="sMA: Phytophthora sojae grown in synthetic medium"
/note="vector: pBK-CW; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 73 a 125 c 106 g 55 t
ORIGIN

Query Match 32.7% Score 17; DB 14; Length 359;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 25 GCTCTGTCGTCGCCG 41
|||||
128 GCTCTGTCGTCGCCG 112

RESULT 27
LOCUS BY076213/c 374 bp mRNA linear EST 06-DEC-2002
DEFINITION BY076213 RIKEN full-length enriched, pooled tissues, adult spleen, etc. Mus musculus cDNA clone K63003015.5', mRNA sequence.
ACCESSION BY076213
VERSION BY076213.1 GI:26177732
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 374)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gotohori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A., Quackenbush, J., Schirimi, L.M., Kanapin, A., Matsuda, H., Batalov, S., Belser, K.W., Blake, J.A., Brad, D., Brusic, V., Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Paven, W.J., Perlea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C.A., Setu, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wysshaw-Borle, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, M., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE
JOURNAL MEDLINE
COMMENT PUBMED
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenhiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsic.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane
, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.,
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,
M., Waki, K., Watabiki, A., Muramatsu, M. and Hayashizaki, Y. Direct
Submission
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

FEATURES

source

location/Qualifiers

```
1. .374
/mol_type="mus musculus"
/db_xref="taxon:10090"
/clone="K630003D15"
/clone_lib="RIKEN full-length enriched, pooled tissues,
adult spleen, etc."
/note="(dev_stage=adult, tissue_type=spleen, sex=male),
(dev_stage=adult, tissue_type=kidney, sex=male),
(dev_stage=adult, tissue_type=testis, sex=male),
(dev_stage=adult, tissue_type=thymus, sex=male),
(dev_stage=adult, tissue_type=heart, sex=male),
(dev_stage=adult, tissue_type=colon, sex=male),
(dev_stage=adult, tissue_type=stomach, sex=male),
(dev_stage=adult, tissue_type=liver, sex=male),
(dev_stage=13 days embryo, tissue_type=whole body, sex=mix
), (dev_stage=16 days embryo, tissue_type=whole body, sex=mix
), (dev_stage=17 days embryo, tissue_type=whole body, sex=mix
), (dev_stage=15 days pregnant, adult, tissue_type=emuron
, sex=female), (dev_stage=10 days neonate, tissue_type=thymus
, sex=mix), (dev_stage=10 days neonate, tissue_type=thymus
, sex=mix), (dev_stage=10 days neonate, tissue_type=heart
, sex=mix)"
```

BASE COUNT

81 a 88 c 144 g 61 t

ORIGIN

Query Match 32.7%; Score 17; DB 13; Length 374;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ATGCGTCTGCTGCTGC 38
|||||
Db 374 ATGCGTCTGCTGCTGC 358

RESULT 28
AU029704/c 404 bp mRNA linear EST 01-APR-2002
LOCUS AU029704 Rice panicle shorter than 3cm Oryza sativa [japonica
DEFINITION cultivar-group) cDNA clone EJ1310_62, mRNA sequence.
ACCESSION AU029704

VERSION AU029704.1 GI:3762952
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriaristidae; Oryzaceae; Oryza.
1 (bases 1 to 404)
AUTHORS Sasaki, T. and Yamamoto, K.
TITLE Rice cDNA from panicle
JOURNAL Unpublished
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>
PROJECT = RGP.
POLY(A)=No.

FEATURES

source

location/Qualifiers

```
1. .404
/mol_type="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="EJ1310_62"
/dev_stage="shorter than 3cm"
/clone_lib="Rice panicle shorter than 3cm"
/note="Organ: panicle"
```

BASE COUNT 91 a 89 c 125 g 97 t 2 others

ORIGIN

Query Match 32.7%; Score 17; DB 9; Length 404;

Best Local Similarity 100.0%; Pred. No. 83;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 CTCTCGTCTGCTGCGGC 42
|||||
Db 162 CTCTCGTCTGCTGCGGC 146

RESULT 29
BF008574/c 406 bp mRNA linear EST 06-OCT-2000
LOCUS BF008574 1613456 Amblyomma americanum larva Lambda Zap Express Amblyomma
DEFINITION americanum cDNA, mRNA sequence.
ACCESSION BF008574
VERSION BF008574.1 GI:10708849
KEYWORDS EST.
SOURCE Amblyomma americanum
ORGANISM Amblyomma americanum
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Parasitiformes; Ixodida; Ixodidae; Amblyomma.
1 (bases 1 to 406)
AUTHORS Hill, C.A. and Gutierrez, J.A.
TITLE Analysis of the expressed genome of the lone star tick, Amblyomma
americanum (Acari: Ixodidae) using an expressed sequence tag
approach
JOURNAL Microb. Comp. Genomics 5 (2), 89-101 (2000) In press
COMMENT Contact: Hill CA
Animal Science Discovery Research
Elanco Animal Health, A Division of Eli Lilly and Company
PO Box 708, 2001 West Main St., Greenfield, IN 46140, USA
Tel: 317 277 0826
Fax: 317 277 4522
Email: HILL_CATHERINE_A@ELILLY.COM.

FEATURES

source

```
1. .406
/mol_type="Amblyomma americanum"
/mol_type="mRNA"
/db_xref="taxon:6943"
/sex="Male, Female"
```

```

/dev stage="larva"
/clone.lib="Amblyomma americanum_larva_lambda_Zap_Express"
/notes="vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      92 a      91 c      118 g      101 t      4 others
ORIGIN

```

Query Match	32.7%	Score 17/	DB 10/	length 406/
Best Local Similarity	100.0%	Pred. No. 83/		
Matches	17/	Conservative	0/	Mismatches 0/
				Indels 0/
				Gaps 0/
Oy	1	ATGCAGTCTCTCAAGC	17	
Db	154	ATGCAGTCTCTCAAGC	138	

RESULT 30	
AM465276/c	
LOCUS	425 bp
DEFINITION	BP230018B10F10 Soates normalized bovine placenta Bos taurus cDNA clone BP230018B10F10 5', mRNA sequence.
ACCESSION	AM465276

ACCESSION	AM4652.10	GI : 7035444
VERSION	AM465276..1	
KEYWORDS	EST.	
SOURCE	Bos taurus	(cow)
ORGANISM	Bos taurus	

REFERENCE
AUTHORS

Lewin, H.A., Soares, M.B., Rebeiz, M., Pardinas, J., Liu, L. and Larson

TITLE BOVINE ESTS
JOURNAL Unpublished
COMMENT Contact: Lewin, H. A.
 W. M. Keck Center for Comparative and Functional Genomics
 University of Illinois at Urbana-Champaign
 340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
 61801, USA
 Tel: 217 333 5998
 Fax: 217 244 5617
 Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National Research Initiative, Animal Genome Resource Grant AG 99-3205-0534 to H. A. Lewin and J. E. Mowack. Base Calling/Quality Scores: PHRED from Washington University Genome Center. Vector Trimmi g: Cross_match from Washington University Genome Center PHAP suite. Sequences submitted are vector free and at least 200 bp in length. PCR primers

FORWARD: TTAATACGACCTCATTATAGCG
BACKWARD: ATTTAACCCCTCATTAAAG
Insert Length: 425 Std Error: 0.00
Plate: BP230018B10 row: F column: 10
Seq primer: AGCGGATTAACAATTTCACACGGA
High quality sequence stop: 425.

```

FEATURES
    source
        Location/Qualifiers
            1..425
                /organism="Bos taurus"
                /mol_type="mRNA"
                /db_xref="taxon:9913"
                /clone="BP23001BBI0F10"
                /sex="female"
                /lab_host="DH10B"
                /clone_id="Soares normalized bovine placenta"
                /note="Organ: placenta; Vector: pRT3pac; Site 1: ECORI;
                Site 2: NotI; The cDNA library was contributed by the
                Soares laboratory and it was constructed and normalized
                as described by Bonaldo, M.F., Lennon, G. and Soares,
                M.B. (1996), Genome Research 6(9) : 791-806. "
BASE COUNT
    93 a 116 c 149 g 67 t

```

Query March 32.7% Score 17; DB 9; length 425;
Best Local Similarity 100.0% Pred. No. 84;

	Matches	17,	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY		22	ATCGCTCTCGTGGCTGC	38						
Dd		335	ATCGCTCTCGTGGCTGC	319						

RESULT	31
AM056821	
LOCUS	
DEFINITION	446 bp mRNA linear EST 27-SEP-1999
ACCESSION	AM056821 Pine Triplex shoot c.l.p library Pinus taeda CDNA clone
VERSION	S756P1.mRNA sequence.
KEYWORDS	AM056821 GI:5929516
SOURCE	EST.
TAXID	Pinus taeda (loblolly pine)

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
<i>Pinus taeda</i>					
Eukaryota: Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;					
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus-					
1 (bases 1 to 446)					
Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.					
The Pine Gene Discovery Project					
Unpublished					
Contact: Rosa Whetten					

FEATUURES
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhe@community.ncsu.edu
Shoot tip library plates ST01-ST50 were prepared and sequenced at
NCsu. Plates ST51-ST99 were prepared and sequenced at IFG/CSUH.
Informatics was done at the Computational Biology Centers (CBC) at
University of Minnesota.
Seq primer: 5' lambdaTriplex2 Sequencing Primer.
location/qualifier

```

SOURCE
1. .446
/locust2/Vectorfields
/organism="Pinus taeda"
/mol_type="mRNA"
/db_xref="taxon:3352"
/clone="ST56F10"
/lab_host="E. coli BM25.8"
/clone_11b="Pine Triplex shoot tip library"
/note="Organ: shoot tips; Vector: Lambda Triplex; Site_1:
Site (A); Site_2: Site (B); Shoot tips (approx. 2 cm from
apex) were collected during the spring, frozen and used
for mRNA isolation. The SMART-PCR method (Clontech) was
used to prepare a library from 1 ug total RNA, using the
Lambda Triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."
BASE COUNT
107 a 137 c 79 g 110 t 13 others
ORIGIN

```

```

Query Match          32.7%; Score 17; DB 9; Length 446;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 17; Conservative 0; Mismatches 0; Gaps 0;

QY      31 GTGCGTGC GGCGCCTCCG 47
        |||||
Gb       80 GTGCGTGC GGCGCCTCCG 96

```

RESULT 32	
AM376507/c	
LOCUS	470 bp mRNA linear EST 04-FEB-2000
DEFINITION	IL3-CT0213-161299-038-F12 CT0213 Homo sapiens cDNA, mRNA sequence.
ACCESSION	U376507
VERSION	AM376507.1 GI:6881168
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 470)
HCGP <http://www.ludwig.org.br/ORESTES>.
The FAPESP/LICR Human Cancer Genome Project
Unpublished
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

FEATURES
source

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=113&ct=113-CT0213-161299-038-F12&ct=1999-12-16&ct=1>)
Seq primer: puc 18 forward
High quality sequence stop: 383.
Location/Qualifiers

BASE COUNT 116 a 119 c 152 g 83 t

ORIGIN

Query Match 32.7%; Score 17; DB 9; Length 470;
Best Local Similarity 100.0%; Pred.No. 85;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ATCGCTCTCGCTCGCTC 38
|||||
276 ATCGCTCTCGCTCGCTC 260

Db

RESULT 33
BQ170403 490 bp mRNA linear EST 26-APR-2002
LOCUS WHE1770.E05 J102T Wheat pre-anthesis spike cDNA library Triticum
DEFINITION aestivum cDNA clone WHE1770.E05.J10, mRNA sequence.
ACCESSION BQ170403.1 GI:20332226
VERSION BQ170403.1
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticaceae; Triticum.
1 (bases 1 to 490)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
The structure and function of the expressed portion of the wheat genomes - pre-anthesis spike cDNA library
Unpublished
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105555773
Fax: 5105555818
Email: anderson@pw.usda.gov

This EST was generated by sequencing from the 3' end of the clone. Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20.
Seq primer: T7 primer.
Location/Qualifiers

FEATURES
source

1..490
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1770.E05.J10"
/tissue_type="Spike_before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/clone_lib="wheat pre-anthesis spike cDNA library"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed. white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the T7 close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 124 a 138 c 117 g 111 t

ORIGIN

Query Match 32.7%; Score 17; DB 13; Length 490;
Best Local Similarity 100.0%; Pred.No. 86;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 CTCTCGTCTGCTCGCGC 42
|||||
345 CTCTCGTCTGCTCGCGC 361

Db

RESULT 34
BE484921/c 505 bp mRNA linear EST 27-MAR-2003
LOCUS BE484921
DEFINITION 171741 BARC SBOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE484921
VERSION BE484921.1 GI:9604454
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
1 (bases 1 to 505)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Sonstegard,T., Capuco,A.V., White,J., Van Tassel,J.C.P., Connor,E.E., Cho,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D. and Quackenbush,J.
Analysis of bovine mammary gland EST and functional annotation of the Bos taurus gene index
Mamm. Genome 13 (7), 373-379 (2002)
22135956
12140684

CONTACT: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@lps1.barc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -mnscore 18 and -mismatch 12 options.
PCR PRIMERS
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGACGACGAC
Place: 135 row: J column: 12
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers

```

source
1. .505
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/issue_type="pooled"
/lab_host="DH10B"
/clone_11b="BARC 5BOV"
/notes="Vector: PCMV SPORT6; Site 1: NotI; Site 2: SalI;
library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."

BASE COUNT      118 a      135 c      175 g      77 t
ORIGIN

Query Match      32.7%; Score 17; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      22 ATCGCTCTGCTGCTGC 38
        |||||
Db      342 ATCGCTCTGCTGCTGC 326

RESULT 35
B0821431      506 bp      mRNA      linear      EST 01-AUG-2002
LOCUS      1030092C04.y1 C. reinhardtii CC-1690, Deflagellation (normalized),
DEFINITION      Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION      B0821431
VERSION      B0821431.1 GI:22072093
KEYWORDS      EST.
ORGANISM      Chlamydomonas reinhardtii
SOURCE      Chlamydomonas reinhardtii
COMMENT      Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
              Chlamydomonadales; Chlamydomonas.
              1 (bases 1 to 506)
              Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre
              P., McDermott, J.P., Shlager, J., Silflow, C. and Stern, D.
              Analyses of the Chlamydomonas reinhardtii Genome: A Model.
              Unicellular System for Analyzing Gene Function and Regulation in
              Vascular Plants. Project: 1030
              Unpublished
              Contact: Charles Hauser
              DCMB Box 91000
              Duke University
              Durham, NC 27708-1000
              Tel: 919 613 8159
              Fax: 919 613 8177
              Email: chauser@duke.edu

FEATURES
source
1. .506
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/db_xref="taxon:3055"
/clone_11b="C. reinhardtii CC-1690, Deflagellation
(normalized), Lambda Zap II"
/notes="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; Deflagellation library, constructed by John Davies
and Jeffrey McDermott, combines cDNAs from CC-1690 cells
which had been re-synthesizing flagella for 15, 30 and 60
min after being deflagellated by pH shock. PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda Zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT      98 a      153 c      168 g      87 t
ORIGIN

Query Match      32.7%; Score 17; DB 13; Length 506;

```

```

Best Local Similarity 100.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      28 CTCGCTGCTGCCGCT 44
        |||||
Db      362 CTCGCTGCTGCCGCT 346

RESULT 36
A1988643      529 bp      mRNA      linear      EST 30-NOV-2001
LOCUS      s05h05.y1 Gm-cl020 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION      Gm-cl020-850 5', mRNA sequence.
ACCESSION      A1988643
VERSION      A1988643.1 GI:5620437
KEYWORDS      EST.
SOURCE      Glycine max (soybean)
ORGANISM      Glycine max
COMMENT      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
              ; eurosoid I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
              Glycine.
              1 (bases 1 to 529)
              Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corryell, V., Khanna
              A., Bolla, B., Maria, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
              Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
              Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck
              R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
              R., Waterston, R. and Wilson, R.
              Public Soybean EST Project
              Unpublished
              Contact: Shoemaker R/Public Soybean EST Project
              Public Soybean EST Project
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              This clone is available through: Resgen, Invitrogen Corp. 2130
              South Memorial Parkway Huntsville, AL 35801 For further information
              call: (800)-533-4363 or contact via email: cdu@resgen.com
              Insert Length: 810 Std Error: 0.00
              Seq primer: -40RP from Gibco
              High quality sequence stop: 429.
              Location/Qualifiers
              1. .529
              /organism="Glycine max"
              /mol_type="mRNA"
              /db_xref="taxon:3847"
              /clone="GENOME SYSTEMS CLONE ID: Gm-cl020-850"
              /issue_type="root nodules of greenhouse grown plants"
              /lab_host="XL10-Gold"
              /clone_11b="Gm-cl020"
              /notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
              XhoI; This cDNA library was constructed from mRNA isolated
              from nodules on the roots of 2.5 month-old Glycine max
              'Williams' plants that were greenhouse grown. The cDNA
              library was prepared using the Stratagene pBluescript II
              SK(+) library construction kit. First strand synthesis was
              performed with 5-methyl dCTP, hence the ligated cDNA was
              hemimethylated. A modification of Stratagene's
              first-strand synthesis primer was used. An anchor'
              nucleotide (v-a, c, or g) was added to the 3' end of the
              primer (GACAGAGAGAGAGAGAGAGACTGCTCGAG(T)18V) to anchor
              the primer at the 5' end of the poly(A) tract. After
              second-strand synthesis, the cDNA ends were filled in with
              cloned Pfu DNA polymerase and size-fractionated with a
              400 bp cutoff, using a SizeSep 400 Spin column from
              Pharmacia. The column eluent was ligated to EcoRI adaptors
              and phosphorylated. The XhoI sites in the cDNA would be
              protected by their hemimethylated status. The cDNA
              constructs were size-fractionated with a 500 bp cutoff,
              using GibcoBRL Life Technologies' cDNA size Predigested

```


vector (pBluescript II SK+) that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=56) and sequence (n=16). This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

BASE COUNT 94 a 157 c 179 g 98 t 1 others

Query Match 32.7%; Score 17; DB 9; Length 529;

Best Local Similarity 100.0%; Pred. No. 88;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 TCTCATGCTCTGCTG 34
|||||
529 TCTCATGCTCTGCTG 513

RESULT 37

AV668034/c

LOCUS AV668034 Bos taurus ovary fetus Bos taurus cDNA clone E10V019D09

DEFINITION 5', mRNA sequence.

ACCESSION AV668034 GI:99327779

VERSION AV668034.1

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 542)

Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.

and Sugimoto,Y.

Establishment of a high throughput EST sequencing system using

poly(A) tail-removed cDNA libraries and determination of 36,000

bovine ESTs

Nucleic Acids Res. 29 (22), E108 (2001)

JOURNAL 11713328

MEDLINE 21570554

PUBMED 11713328

COMMENT Contact: Yoshikazu Sugimoto

Animal Genetics Division

Shirakawa Institute of Animal Genetics

Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazusugi@cocoa.ocn.ne.jp

Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

Location/Qualifiers

1..542

/organism="Bos taurus"

/mol_type="mRNA"

/db_xref="taxon:9913"

/clone="E10V019D09"

/tissue_type="ovary"

/dev_stage="Fetus"

/lab_host="DH10B"

/clone_lib="Bos taurus ovary fetus"

/note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A

was deleted from a NotI site"

BASE COUNT 121 a 153 c 187 g 81 t

ORIGIN

Query Match 32.7%; Score 17; DB 9; Length 542;

Best Local Similarity 100.0%; Pred. No. 88;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 22 ATGCGTCTGCTGCTG 38

|||||

375 ATGCGTCTGCTGCTG 359

RESULT 38

A1542489 544 bp mRNA linear EST 23-APR-2001

LOCUS A1542489/c

DEFINITION SD08831.5prime SD Drosophila melanogaster Schneider L2 cell culture

POT2 Drosophila melanogaster cDNA clone SD08831.5 similar to

AJ010298: Drosophila melanogaster retrotransposon-like element,

mRNA sequence.

ACCESSION A1542489

VERSION A1542489.2 GI:13771703

KEYWORDS EST.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 544)

Harvey,D., Broksstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,

Lewis,S. and Rubin,G.M.

BDGP/HMWI Drosophila EST Project

Unpublished

On Mar 19, 1999 this sequence version replaced gi:4459862.

Other ESTs: SD08831.3prime

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

hit genomic AEO03350: arm:U (4635408,4651408) estimated-cyto.?:

04/13/2001

Plate: SD.88 row: C column: 7

High quality sequence stop: 459

POLYA=NO.

Location/Qualifiers

1..544

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="SD08831"

/lab_host="DH5-alpha"

/clone_lib="SD Drosophila melanogaster Schneider L2 cell

culture POT2"

/note="Vector: POT2; Site_1: EcoRI; Site_2: XhoI; Sized

fractionated cDNAs were directly ligated into POT2.

Plasmid cDNA library."

BASE COUNT 119 a 148 c 171 g 106 t

ORIGIN

Query Match 32.7%; Score 17; DB 9; Length 544;

Best Local Similarity 100.0%; Pred. No. 88;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 CATGCTCTGCTGCTG 37

|||||

471 CATGCTCTGCTGCTG 455

RESULT 39

AV597173 546 bp mRNA linear EST 27-NOV-2001

LOCUS AV597173/c

DEFINITION AV597173 Bos taurus cartilage fetus Bos taurus cDNA clone

E1CA030D01.5', mRNA sequence.

ACCESSION AV597173

VERSION AV597173

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 546)

Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.

and Sugimoto,Y.

TITLE Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs

JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)

MEDLINE 21570554

PUBMED 11713328

COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
OdaKura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5725
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES

source

1..546
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1CA030D01"
/tissue_type="cartilage"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_1lb="Bos taurus cartilage fetus"
/note="Vector: pZLI; Site_1: SalI; Site_2: NotI; Poly A was deleted from a NotI site"

BASE COUNT 126 a 148 c 180 g 89 t 3 others

ORIGIN

Query Match 32.7%; Score 17; DB 9; Length 546;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ATCGCTCTCGTCGCTGC 38
|||||
346 ATCGCTCTCGTCGCTGC 330

RESULT 40

LOCUS AV603191 556 bp mRNA linear EST 27-NOV-2001

DEFINITION AV603191 Bos taurus kidney fetus Bos taurus cDNA clone E1K1014F12 5', mRNA sequence.

ACCESSION AV603191

VERSION AV603191.1 GI:9725517

KEYWORDS EST.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 556)
Takesuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H. and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs

TITLE

JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)

MEDLINE 21570554

PUBMED 11713328

COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
OdaKura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5725
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES

source 1..556
/organism="Bos taurus"

/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1K1014F12"
/tissue_type="kidney"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_1lb="Bos taurus kidney fetus"
/note="Vector: pZLI; Site_1: SalI; Site_2: NotI; Poly A was deleted from a NotI site"

BASE COUNT 127 a 150 c 187 g 89 t 3 others

ORIGIN

Query Match 32.7%; Score 17; DB 9; Length 556;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 ATCGCTCTCGTCGCTGC 38
|||||
359 ATCGCTCTCGTCGCTGC 343

Search completed: November 13, 2003, 10:59:31
Job time : 2025 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 13, 2003, 10:59:37 ; Search time 218 Seconds
(without alignments)
222.889 Million cell updates/sec

Title: US-10-081-935-2
Perfect score: 82
Sequence: 1 MGFHALIALVAGLASA 18

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Xgapop 6.0 , Ygapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=x1h
-O=/cgn2.1/USPRO.spool/US10081935/funat.12112003.144406.3612/app.query.fasta.1.199
-DB=N_Geneseq.13Jun03 -OMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=Pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081935 @CGN.1.1.0 @runat.12112003.144406.3612 -NCPU=6 -ICPU=3
-NO_MMAP -JARGOUTERY -NEG_SCORES=0 -MAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
N_Geneseq.13Jun03:*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	61.0	450	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
2	47	57.3	5888	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
3	47	57.3	4403765	22	Human ORFX polynucleotide sequence SEQ ID NO:10167.
4	47	57.3	4411529	22	Human ORFX polynucleotide sequence SEQ ID NO:10167.
5	46	56.1	7788	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
6	46	56.1	10323	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
7	44	53.7	1751	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
8	44	53.7	1844	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
9	44	53.7	1947	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
10	44	53.7	4751	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
11	44	53.7	4754	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
12	44	53.7	4790	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
13	44	53.7	10091	19	Human ORFX polynucleotide sequence SEQ ID NO:10167.
14	43	52.4	549	21	Human ORFX polynucleotide sequence SEQ ID NO:10167.
15	43	52.4	549	21	Human ORFX polynucleotide sequence SEQ ID NO:10167.
16	43	52.4	1554	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
17	43	52.4	1758	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
18	43	52.4	6783	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
19	43	52.4	10035	25	Human ORFX polynucleotide sequence SEQ ID NO:10167.
20	43	52.4	11115	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
21	43	52.4	11220	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
22	43	52.4	56609	21	Human ORFX polynucleotide sequence SEQ ID NO:10167.
23	43	52.4	349980	21	Human ORFX polynucleotide sequence SEQ ID NO:10167.
24	43	52.4	349980	21	Human ORFX polynucleotide sequence SEQ ID NO:10167.
25	43	52.4	1437668	21	Human ORFX polynucleotide sequence SEQ ID NO:10167.
26	42	51.2	498	22	Human ORFX polynucleotide sequence SEQ ID NO:10167.
27	42	51.2	504	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
28	42	51.2	806	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
29	42	51.2	881	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
30	42	51.2	951	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
31	42	51.2	1031	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
32	42	51.2	1201	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
33	42	51.2	1281	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
34	42	51.2	1474	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
35	42	51.2	1569	23	Human ORFX polynucleotide sequence SEQ ID NO:10167.
36	42	51.2	1714	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
37	42	51.2	1776	15	Human ORFX polynucleotide sequence SEQ ID NO:10167.
38	42	51.2	1824	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
39	42	51.2	1874	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
40	42	51.2	2003	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
41	42	51.2	2027	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
42	42	51.2	2053	24	Human ORFX polynucleotide sequence SEQ ID NO:10167.
43	42	51.2	2211	15	Human ORFX polynucleotide sequence SEQ ID NO:10167.
44	42	51.2	2211	15	Human ORFX polynucleotide sequence SEQ ID NO:10167.
45	42	51.2	2211	21	Human ORFX polynucleotide sequence SEQ ID NO:10167.

ALIGNMENTS

RESULT 1
ABN20845 standard; CDNA; 450 BP.
ABN20845:
24-JUN-2002 (first entry)
Human ORFX polynucleotide sequence SEQ ID NO:10167.
Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
degenerative disorder; osteoarthritis; neurodegenerative disorder;
cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
hypertension; hypothyroidism; cholesterol ester storage disease;
immune deficiency; immune disorder; infectious disease;
autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;


```

RESULT 3
AA199683
ID AA199683 standard; DNA: 4403765 BP.
XX AC
XX AA199683;
XX DT 15-JAN-2002 (first entry)
XX DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX OS Mycobacterium tuberculosis.
XX PN US6294328-B1.
XX PD 25-SEP-2001.
XX PF 24-JUN-1998; 98US-0103840.
XX PR 24-JUN-1998; 98US-0103840.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX WPI; 2001-647261/74.
XX PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
XX PS
XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX CC
XX The invention relates to evaluating strain variation within and between
XX different populations of the tuberculosis bacterial pathogen,
XX Mycobacterium tuberculosis or related Mycobacterium by determining the
XX nucleotide sequence of the first strain at positions in the complete
XX sequence of the genome that correspond to positions that differ in the
XX nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
XX H37Rv (AA199682). The method is useful for evaluating strain variation of
XX M. tuberculosis and has valuable application in the fields of
XX tuberculosis genetics, epidemiology, patient treatment and epidemic
XX monitoring.
XX CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from USPTO
XX at seqdata.uspto.gov/sequence.html?DocID=62994328B1.
XX SO Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T, 189 other.

Alignment Scores:
Pred. No.: 8,89e+05 Length: 4403765
Score: 47.00 Matches: 9
Percent Similarity: 92.86% Conservative: 4
Best Local Similarity: 64.29% Mismatches: 1
Query Match: 57.32% Indels: 0
DB: 22 Gaps: 0

US-10-081-935-2 (1-18) x AA199683 (1-4403765)
QY 3 PhSeRHiaLeuIlleAlaLeuValAlaAlaGlyLeuAla 16
ID AA199682 standard; DNA: 4411529 BP.
XX AC
XX AA199682;
XX DT 15-JAN-2002 (first entry)

```

```

XX DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX OS Mycobacterium tuberculosis.
XX PV US6294328-B1.
XX PD 25-SEP-2001.
XX PF 24-JUN-1998; 98US-0103840.
XX PR 24-JUN-1998; 98US-0103840.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX DR WPI; 2001-647261/74.
XX PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AI199683) and
CC H37Rv (AI199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docid=6294328B1.
XX SO Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

Alignment Scores:
Pred. No.:      8.9e+05      Length:      4411529
Score:          47.00        Matches:       9
Percent Similarity: 92.86%    Conservative:   4
Best Local Similarity: 64.29%  Mismatches:    1
Query Matchn:     57.32%     Indels:         0
DB:              22         Gaps:           0

US-10-081-935-2 (1-18) x AA199682 (1-4411529)

QY          3 PheserHtAlaleuIlleAlaleuVaIAAalagIleuAa 16
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1219072 TTCGGCAGCGCCCTGTGTCATCTTGCCTGCCGTAATTGCGC 1219113

RESULT 5
AA584289
ID AA584289 standard; cDNA; 7788 BP.
XX AC AA584289;
XX DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #20093.
XX KM Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
```

Mouse: jachamr, 2000

Scores: 1313 A; 1004 C; 980 G; 1457 T; 0 other.

Pred. No.: 833 Length: 4754
Score: 44.00 Matches: 10
Percent Similarity: 80.00% Conservative: 2
Best Local Similarity: 66.67% Mismatches: 3
Query Match: 53.66% Indels: 0
DB: 23 Gaps: 0

US-10-081-935-2 (1-18) x ABL17908 (1-4754)

OY 4 SerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
DB 2107 GCACATGCCCGCCATTGCCCTTACTTGCAGCCCAAGCCCAAGCC 2151

RESULT 12

ABL12108
ID ABL12108 standard; cDNA; 4790 BP.

XX ABL12108;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 30806.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB68005.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

PS Claim 1; SEQ ID NO 30806; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB5737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4790 BP; 1324 A; 1012 C; 986 G; 1468 T; 0 other;

US-10-081-935-2 (1-18) x ABL12108 (1-4790)

OY 4 SerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
DB 2107 GCACATGCCCGCCATTGCCCTTACTTGCAGCCCAAGCCCAAGCC 2151

RESULT 13

AAV69642
ID AAV69642 standard; DNA; 10091 BP.

XX AAV69642;

XX 02-FEB-1999 (first entry)

DE Drosophila fat facets related Y (DFFRY) gene sequence.

XX Non-recombining region; human; Y chromosome; X homologue; testis;

KW infertility; sperm; Drosophila fat facets related Y; inhibitor;

KW gene alteration; DFFRY; ss.

XX Homo sapiens.

XX WO9846747-A2.

XX 22-OCT-1998.

XX 10-APR-1998; 98WO-US07115.

XX 11-APR-1997; 97US-0041877.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX Lahn BT, Page DC;

XX WPI; 1998-568729/48.

DR P-PSDB; AAW61512.

XX Novel genes in the non-combining region of Y chromosome - useful to

PT diagnose if male infertility or reduced sperm count has a genetic

PT basis

XX Claim 7; Fig 7A-D; 54pp; English.

XX This represents the nucleotide sequence of Drosophila fat facets related

CC Y (DFFRY) gene. The invention relates to genes occurring on the non-

CC recombining region of the human Y chromosome. The sequences fall

CC into two classes: (1) X-homologous DNA which are expressed in many

CC organs, having functional X homologues and (2) testis-specific DNA

CC sequences. Y chromosomal DNA from males with known conditions such as

CC infertility and reduced sperm count can be assessed using the invention

CC to determine whether the condition is associated with or caused by the

CC occurrence of the gene or gene alteration. Candidate inhibitors of the

CC enzymatic activity of the genes can be assessed using in vitro assays.

XX Sequence 10091 BP; 3208 A; 1797 C; 2103 G; 2983 T; 0 other;

US-10-081-935-2 (1-18) x AAV69642 (1-10091)

OY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSer 17

DB 3483 CTTCAACAATAATCATGCTTGTACTTGTGACGAGAAACCTTGCAAC 3533

RESULT 14

AAAC94394

ID AAC94394 standard; cDNA; 549 BP.

AC AAC94394;
 XX 19-FEB-2001 (first entry)
 XX
 DE Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:889.
 XX
 KM Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;
 XX vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.
 XX
 OS Ctenocephalides felis.
 XX
 PN WO200061621-A2.
 XX
 PD 19-OCT-2000.
 XX
 PF 07-APR-2000; 2000WO-US09437.
 XX
 PR 09-APR-1999; 99US-0128704.
 XX
 PA (HESK-) HESKA CORP.
 XX
 PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
 XX
 DR WPI; 2000-656323/63.
 XX
 PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
 PT acids useful for the prevention, diagnosis and treatment of flea
 PT infestations -
 XX
 PS Claim 26; Page 538; 964pp; English.
 XX
 CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
 CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
 CC or head and nerve cord (HNC) tissue. The invention also relates to the
 CC encoded proteins. The invention additionally encompasses expression
 CC constructs, recombinant viruses and recombinant cells comprising the
 CC nucleic acids of the invention, recombinant production of the proteins,
 CC antibodies against the proteins, a method of identifying inhibitors of
 CC the proteins, and compositions comprising the inhibitors for
 CC administration to an animal. The nucleic acids, and the proteins they
 CC encode may be used in the prevention, treatment and diagnosis of diseases
 CC associated with flea infestations. For example, the nucleic acids may be
 CC used to produce an HMT or HNC protein according to standard recombinant
 CC DNA methodology by inserting the nucleic acids into a host cell and
 CC culturing the cell to express the protein. The HMT and HNC nucleic acids
 CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
 CC and quantitate the presence of cat flea or other homologous nucleic acid
 CC sequences in samples. They may also be used to study the expression and
 CC function of the proteins and their role in metabolism. The HMT and HNC
 CC proteins may be used as antigens in the production of specific
 CC antibodies, and in assays to identify modulators (agonists and
 CC antagonists) of HMT and/or HNC protein expression and activity. The
 CC anti-HMT/HNC protein antibodies and antagonists may also be used to
 CC downregulate protein expression and activity. The antibodies may also be
 CC used as diagnostic agents for detecting the presence of flea polypeptides
 CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
 CC present sequence represents a cat flea HNC cDNA of the invention.
 XX
 XX
 SQ Sequence 549 BP; 158 A; 126 C; 137 G; 128 T; 0 other;
 XX
 XX
 Alignment Scores:
 Pred. No.: 91.4 Length: 549
 Score: 43.00 Matches: 9
 Percent Similarity: 72.22% Conservative: 4
 Best Local Similarity: 50.00% Mismatches: 5
 Query Match: 52.44% Indels: 0
 DB: 21 Gaps: 0
 US-10-081-935-2 (1-18) x AAC94394 (1-549)
 OY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerIle 18
 DB 21 ATGAATTCGCAGTAGCAATTTTGGCCCTGCTGTGCTTGGATCAGACT 74

RESULT 15
 AAC95143
 ID AAC95143 standard; cDNA; 549 BP.
 XX
 AC AAC95143;
 XX
 DE 19-FEB-2001 (first entry)
 XX
 DE Cat flea head and nerve cord (HNC) cDNA, SEQ ID NO:1638.
 XX
 KM Cat flea; head and nerve cord nucleic acid; HNC; flea infestation;
 XX vaccine; antiparasitic; therapeutic target; diagnosis; detection; ss.
 XX
 OS Ctenocephalides felis.
 XX
 PN WO200061621-A2.
 XX
 PD 19-OCT-2000.
 XX
 PF 07-APR-2000; 2000WO-US09437.
 XX
 PR 09-APR-1999; 99US-0128704.
 XX
 PA (HESK-) HESKA CORP.
 XX
 PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
 XX
 DR WPI; 2000-656323/63.
 XX
 PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
 PT acids useful for the prevention, diagnosis and treatment of flea
 PT infestations -
 XX
 PS Claim 26; Page 793; 964pp; English.
 XX
 CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
 CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
 CC or head and nerve cord (HNC) tissue. The invention also relates to the
 CC encoded proteins. The invention additionally encompasses expression
 CC constructs, recombinant viruses and recombinant cells comprising the
 CC nucleic acids of the invention, recombinant production of the proteins,
 CC antibodies against the proteins, a method of identifying inhibitors of
 CC the proteins, and compositions comprising the inhibitors for
 CC administration to an animal. The nucleic acids, and the proteins they
 CC encode may be used in the prevention, treatment and diagnosis of diseases
 CC associated with flea infestations. For example, the nucleic acids may be
 CC used to produce an HMT or HNC protein according to standard recombinant
 CC DNA methodology by inserting the nucleic acids into a host cell and
 CC culturing the cell to express the protein. The HMT and HNC nucleic acids
 CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
 CC and quantitate the presence of cat flea or other homologous nucleic acid
 CC sequences in samples. They may also be used to study the expression and
 CC function of the proteins and their role in metabolism. The HMT and HNC
 CC proteins may be used as antigens in the production of specific
 CC antibodies, and in assays to identify modulators (agonists and
 CC antagonists) of HMT and/or HNC protein expression and activity. The
 CC anti-HMT/HNC protein antibodies and antagonists may also be used to
 CC downregulate protein expression and activity. The antibodies may also be
 CC used as diagnostic agents for detecting the presence of flea polypeptides
 CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
 CC present sequence represents a cat flea HNC cDNA of the invention.
 XX
 XX
 SQ Sequence 549 BP; 158 A; 126 C; 137 G; 128 T; 0 other;
 XX
 XX
 Alignment Scores:
 Pred. No.: 91.4 Length: 549
 Score: 43.00 Matches: 9
 Percent Similarity: 72.22% Conservative: 4
 Best Local Similarity: 50.00% Mismatches: 5
 Query Match: 52.44% Indels: 0
 DB: 21 Gaps: 0

US-10-081-935-2 (1-18) x AAC95143 (1-549)

OY 1 MetGlnPheSerIstAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
 DB 21 ATGAAATTCGACGTAGCAATTTGGCGCTGCTGCTGTTGGCATCAGCT 74

RESULT 16

ID AB266799 standard; DNA; 1554 BP.

AC AB266799;

DT 21-MAR-2003 (first entry)

DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 260.

KW Orthosomycin; biosynthesis; evernimicin; avilamycin; gene; ds.

OS Micromonospora carbonacea africana.

XX WO200279505-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-CA00432.

XX 28-MAR-2001; 2001US-279095P.

XX 30-MAR-2001; 2001US-279709P.

XX 20-APR-2001; 2001US-285214P.

XX (ECOP-) ECOP1A BIOSCIENCES INC.

XX Farnet CM, Zazopoulos E, Staffa A;

XX WPI: 2003-058435/05.

XX P-PSDB: ABP99336.

XX Identifying orthosomycin biosynthetic gene, gene fragment or gene cluster, by detecting presence of nucleic acid sequence corresponding to 17 of flamhamycins protein families -

XX Example 1; Page 409; 511pp; English.

XX The invention relates to identifying orthosomycin biosynthetic genes and its fragment/gene cluster (AB266670-AB266681), comprising detecting

XX the presence of a nucleic acid sequence coding for a polypeptide (ABP99207-ABP99362). The method is useful for identifying an orthosomycin

XX biosynthetic gene, gene fragment or gene cluster, especially an evenimicin-type or avilamycin-type orthosomycin biosynthetic gene,

XX gene fragment or gene cluster. The method is useful for detecting the presence of any organism that contains DNA for the production of

XX orthosomycins (both evenimicin-type orthosomycins and avilamycin-type orthosomycins) regardless of the level at which genes for orthosomycin

XX production are expressed by the organism or the amount of orthosomycin produced by the organism. This allows for the detection of new

XX orthosomycin natural products, not produced by the organism.

XX Sequence 1554 BP; 267 A; 615 C; 474 G; 198 T; 0 other;

Alignment Scores:

Pred. No.: 326 Length: 1554

Score: 43.00 Matches: 12

Percent Similarity: 63.64% Conservative: 2

Best Local Similarity: 54.55% Mismatches: 2

Query Match: 52.44% Indels: 6

DB: 25 Gaps: 1

US-10-081-935-2 (1-18) x AB266799 (1-1554)

OY 1 MetGlnPheSerIstAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 14
 DB 1 ATGAAATTCGACGTAGCAATTTGGCGCTGCTGCTGTTGGCATCAGCT 60

OY 15 LeuAla 16
 DB 61 CTGCCC 66

RESULT 17
 ID AAS90583 standard; cDNA; 1758 BP.

AC AAS90583;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #26387.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB: ABG26396.

XX New isolated polynucleotide and encoded polypeptides, useful in diagnosis, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits and to assess biodiversity -

XX Claim 1; SEQ ID NO 26387; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The

XX polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques

XX to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or

XX quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical

XX imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity.

XX The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and

XX amino acid sequences. AAS64197-AAS94564 represent novel human

XX diagnostic coding sequences of the invention.

XX Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

Alignment Scores:

Pred. No.: 379 Length: 1758

Score: 43.00 Matches: 8

Percent Similarity: 80.00% Conservative: 4

Best Local Similarity: 53.33% Mismatches: 3

Query Match: 52.44% Indels: 0

DB: 23 Gaps: 0

US-10-081-935-2 (1-18) x AAS90583 (1-1758)

Qy 3 PheserhisalaleuilealaLeuValAlaAlaGlyLeuAlaSer 17
 Db 416 TACCAGCATTTATTATTAGTTAGTCAGGCGAGGCTAGCTCC 460

RESULT 18
 ABL03546
 ID ABL03546 standard; cDNA; 6783 BP.
 XX
 AC ABL03546;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5120.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 XX
 OS pharmaceutical; gene; ss.
 XX
 PN Drosophila melanogaster.
 XX
 MO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 DR P-PSDB; ABB59443.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 5120; 21pp + Sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 CC
 XX
 XX Sequence 6783 BP; 2354 A; 1063 C; 1140 G; 2226 T; 0 other;

Alignment Scores:

Pred. No.:	1.98e+03	Length:	6783
Score:	43.00	Matches:	9
Percent Similarity:	80.00%	Conservative:	3
Best Local Similarity:	60.00%	Mismatches:	3
Query Match:	52.44%	Indels:	0
DB:	23	Gaps:	0

US-10-081-935-2 (1-18) x ABL03546 (1-6783)

Qy 3 PheserhisalaleuilealaLeuValAlaAlaGlyLeuAlaSer 17
 Db 5989 TTTTCAATCTTATTATTAACACTCTGTCGACGAGTTGGCTTCT 6033
 RESULT 19

ABZ66813/C
 ID ABZ66813 standard; DNA; 10035 BP.
 XX
 AC ABZ66813;
 XX
 DT 21-MAR-2003 (first entry)
 XX
 DE Orthosomycin biosynthetic gene cluster SEQ ID NO 282.
 XX
 KM Orthosomycin; biosynthesis; evernimycin; avilamycin; gene; ds.
 XX
 OS Micromonospora carbonacea aurantiaca.
 XX
 PN WO200279505-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 28-MAR-2002; 2002WO-CA00432.
 XX
 PR 28-MAR-2001; 2001US-279095P.
 XX
 PR 30-MAR-2001; 2001US-279709P.
 XX
 PR 20-APR-2001; 2001US-285214P.
 XX
 PA (ECOP-) ECOPIA BIOSCIENCES INC.
 XX
 PI Farnet CM, Zazopoulos E, Staffa A;
 XX
 DR WPI; 2003-058435/05.
 XX
 PT Identifying orthosomycin biosynthetic gene, gene fragment or gene
 PT cluster, by detecting presence of nucleic acid sequence corresponding
 PT to 17 of flambamycins protein families -
 XX
 PS Example 1; Page 506-511; 511pp; English.
 XX
 CC The invention relates to identifying orthosomycin biosynthetic genes and
 CC its fragment/gene cluster (ABZ66670-ABZ66813), comprising detecting
 CC the presence of a nucleic acid sequence coding for a polypeptide
 CC (ABP99207-ABP99362). The method is useful for identifying an orthosomycin
 CC biosynthetic gene, gene fragment or gene cluster, especially an
 CC evernimycin-type or avilamycin-type orthosomycin biosynthetic gene,
 CC gene fragment or gene cluster. The method is useful for detecting the
 CC presence of any organism that contains DNA for the production of
 CC orthosomycins (both evernimycin-type orthosomycins and avilamycin-type
 CC orthosomycins) regardless of the level at which genes for orthosomycin
 CC production are expressed by the organism or the amount of orthosomycin
 CC produced by the organism. This allows for the detection of new
 CC orthosomycin natural products, not produced by the organism.
 XX
 XX Sequence 10035 BP; 1396 A; 3516 C; 3657 G; 1466 T; 0 other;

Alignment Scores:

Pred. No.:	3.19e+03	Length:	10035
Score:	43.00	Matches:	12
Percent Similarity:	63.64%	Conservative:	2
Best Local Similarity:	54.55%	Mismatches:	2
Query Match:	52.44%	Indels:	6
DB:	25	Gaps:	1

US-10-081-935-2 (1-18) x ABZ66813 (1-10035)

Qy 1 MetGlnPheserhisalaleuileala-----leuValAlaAlaGly 14
 Db 7811 ATGAGACATCTGATGCCCTGCGCGGACACACCGACCGACTGTCGCGCGCGC 7752

Qy 15 leuAla 16
 Db 7751 CTGGCC 7746

RESULT 20
 ID ABL50562/c
 XX ABL50562 standard; DNA; 11115 BP.

AC ABL50562;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Micromonospora carbonacea evernimycin locus nucleotide contig 8.
 XX
 KM Micromonospora carbonacea; antibiotic; evernimycin; biosynthesis;
 XX gene cluster; genetic manipulation; contig; gene; ds.
 XX
 OS Micromonospora carbonacea.
 XX
 PN WO200155180-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-CA00128.
 XX
 PR 27-JAN-2000; 2000US-0177711.
 XX
 PA (ECOP-) ECOPIA BIOSCIENCES INC.
 PA (FARN/) FARNET C.
 PI Staffa A, Zazopoulos E, Mercure S, Nowacki P;
 XX
 XX WPI: 2001-476185/51
 DR P-PSDB; ABB06922, ABB06923, ABB06924, ABB06925, ABB06926, ABB06927,
 DR ABB06928, ABB06929, ABB06930.
 XX
 PT Novel isolated gene cluster encoding polypeptides involved in
 PT evernimycin biosynthesis useful for construction of evernimycin
 PT overproducing strains, and to allow chemical modifications of
 PT evernimycin to enhance certain properties -
 XX
 XX Claim 1; Fig 1; 181pp; English.
 PS
 XX ABL50562 represent contigs 1 to 8 from the Micromonospora
 CC carbonacea evernimycin biosynthetic locus gene cluster. The contigs
 CC encode the protein sequences designated ORF (open reading frame) 1 to
 CC 49, given in ABB06881 to ABB06930. The gene cluster is useful for the
 CC construction of the evernimycin antibiotic in overproducing strains,
 CC and to allow chemical modifications of evernimycin to enhance certain
 CC properties via genetic manipulation or combinatorial biosynthesis. The
 CC gene cluster can be used to produce genetic systems and genes encoding
 CC novel enzyme activities, and avoid the problems of low yield and quality
 CC of evernimycin produced by chemical synthesis.
 CC
 SQ Sequence 11115 BP; 1521 A; 3924 C; 4057 G; 1613 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 3.62e+03 Length: 11115
 Score: 43.00 Matches: 12
 Percent Similarity: 63.64% Conservative: 2
 Best Local Similarity: 54.55% Mismatches: 2
 Query Match: 52.44% Indels: 6
 DB: 23 Gaps: 1
 US-10-081-935-2 (1-18) x ABL50562 (1-11115)
 OY 1 MercinPheserHisAlaLeuVala-----leuValaAlaacy 14
 DB 7824 ATGAGACATCTCTATGCTCTGCGCTCGGCACACCGACCGACCTGTGCGCGCGCGC 7765
 OY 15 leuAla 16
 DB 7764 CTGGCC 7759
 RESULT 21
 ID AAS87868 standard; cDNA; 11220 BP.
 AC AAS87868;
 XX
 XX 13-FEB-2002 (first entry)
 DT

XX
 DE DNA encoding novel human diagnostic protein #23672.
 XX
 XX Human; Chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI: 2001-639362/73.
 DR P-PSDB; ABG23681.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 1; SEQ ID NO 23672; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probe,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 11220 BP; 4082 A; 2495 C; 2204 G; 2437 T; 2 other;
 XX
 Alignment Scores:
 Pred. No.: 3.66e+03 Length: 11220
 Score: 43.00 Matches: 8
 Percent Similarity: 80.00% Conservative: 4
 Best Local Similarity: 53.33% Mismatches: 3
 Query Match: 52.44% Indels: 0
 DB: 23 Gaps: 0
 US-10-081-935-2 (1-18) x AAS87868 (1-11220)
 OY 3 PheSerHisAlaLeuValaLeuValaAlaacyLeuAlaSer 17
 DB 7210 TACCAGCATTTATTATTAAGTTAGTGAGCGGAGGAGCTCC 7254
 RESULT 22
 ID AAA81459/C.
 XX
 XX AAA81459 standard; DNA; 56609 BP.
 XX

AC AAA81459;
 XX 04-DEC-2000 (first entry)
 XX N. meningitidis partial DNA sequence gnm_7 SFO ID NO:7.
 DE
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 XX Meningococcus B; MenB; ds.
 XX
 XX Neisseria meningitidis.
 OS
 XX WO20002430-A2.
 PN
 XX 20-APR-2000.
 PD
 XX 08-OCT-1999; 99WO-US23573.
 PF
 XX 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 XX (CHIR) CHIRON CORP.
 PA
 XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 XX WPI: 2000-318079/27.
 DR
 XX
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisseria infections, for example, N.gonorrhoea -
 PT
 XX
 XX Claim 7: Page 303-320; 1760pp; English.
 PS
 XX
 XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 CC
 XX
 XX Sequence 56609 BP; 14040 A; 15839 C; 14415 G; 12313 T; 2 other;
 SQ
 Alignment Scores:
 Pred. No.: 2.65e+04 Length: 56609
 Score: 43.00 Matches: 9
 Percent Similarity: 92.31% Conservative: 3
 Best Local Similarity: 69.23% Mismatches: 1
 Query Match: 52.44% Indels: 0
 DB: 21 Gaps: 0
 US-10-081-935-2 (1-18) x AAA81459 (1-56609)

QY 6 AAlaenUlaAlaenValAlaAagUlyLeuAlaSerAla 18
 DB 8690 GCCTTCATCTGATGCTCGGCTCTCGCGGCT 8652
 RESULT 23
 AAF21609
 ID AAF21609 standard; DNA; 349980 BP.
 XX
 AC AAF21609;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:110.
 DE
 XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 XX ds.
 XX
 XX Neisseria meningitidis.
 OS
 XX WO200066791-A1.
 PN
 XX 09-NOV-2000.
 PD
 XX 08-MAR-2000; 2000MO-US05928.
 PF
 XX 30-APR-1999; 99US-0132068.
 PR 08-OCT-1999; 99WO-US23573.
 PR 28-FEB-2000; 2000GB-0004695.
 XX
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
 PI Frazer CM, Grandi G;
 XX
 XX WPI: 2000-647603/62.
 DR
 XX
 XX Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisserial infections -
 PT
 XX
 XX Claim 7; Appendix A; 692pp; English.
 F:
 XX
 XX The present invention describes the full length genome of
 CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own it was split into 8
 CC sequences which overlap each other at the beginning and end of each
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
 CC AAF21606 represent PCR primers which are used in the exemplification of
 CC the present invention. The NMB genome and fragments from it have
 CC antibacterial activity, and can be used in vaccines and gene therapy.
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
 CC proteins can be used in compositions for treating or preventing infection
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
 CC presence of Neisserial bacteria or of antibodies raised to Neisserial
 CC bacteria. Computers, computer memory, computer storage medium or computer
 CC databases can be used in a search to identify open reading frames (ORFs)
 CC or coding sequences within the NMB genome. The DNA sequences provide
 CC further opportunities to find antigenic or immunogenic proteins which are
 CC more effective in vaccines than the outer membrane proteins currently
 CC used.
 CC
 XX
 XX Sequence 349980 BP; 81351 A; 86755 C; 95584 G; 86290 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 2.4e+05 Length: 349980
 Score: 43.00 Matches: 9
 Percent Similarity: 92.31% Conservative: 3

```

Best Local Similarity: 69.23% Mismatches: 1
Query Match: 52.44% Indels: 0
DB: 21 Gaps: 0

US-10-081-935-2 (1-18) x AAF21609 (1-349980)

OY 6 A1aleuilealaleuVal1Ala1aG1yleuA1aSerA1a 18
Db 317275 GCCTTGCTCATTCGATTGTCGCGGCTCTCGCTCGGCT 317313

RESULT 24
AAF21610
ID AAF21610 standard; DNA; 349980 BP.
XX
AC AAF21610;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:111.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KM ds.
XX
OS Neisseria meningitidis.
XX
PN WO20006791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
PI Pizsa M, Hickey E, Peterson J, Tetteijn H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
XX WPI; 2000-647603/62.
XX
XX Neisseria meningitidis B full length genome sequence and open reading
XX frames are used to detect, treat and prevent Neisserial infections -
XX
XX Claim 7; Appendix A; 692pp; English.
XX
XX The present invention describes the full length genome of
XX Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
XX to AAF21613 represent fragments of the NMB genomic sequence, as the
XX sequence was too long to go in a record on its own it was split into 8
XX sequences which overlap each other at the beginning and end of each
XX the beginning of AAF21607, the last 49980 bp of AAF21544 is repeated at
XX the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
XX Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
XX AAF21606 represent PCR primers which are used in the exemplification of
XX the present invention. The NMB genome and fragments from it have
XX antibacterial activity, and can be used in vaccines and gene therapy.
XX Neisseria nucleic acids, proteins and/or antibodies which binds to the
XX proteins can be used in compositions for treating or preventing infection
XX due to Neisserial bacteria or as a diagnostic reagent for detecting the
XX presence of Neisserial bacteria or of antibodies raised to Neisserial
XX bacteria. Computers, computer memory, computer storage medium or computer
XX databases can be used in a search to identify open reading frames (ORFs)
XX or coding sequences within the NMB genome. The DNA sequences provide
XX further opportunities to find antigenic or immunogenic proteins which are
XX more effective in vaccines than the outer membrane proteins currently
XX used.

```

```

SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;
Alignment Scores:
Pred. No.: 2.4e+05
Score: 43.00 Length: 349980
Percent Similarity: 92.31%
Best Local Similarity: 69.23% Mismatches: 3
Query Match: 52.44% Indels: 1
DB: 21 Gaps: 0

US-10-081-935-2 (1-18) x AAF21610 (1-349980)

OY 6 A1aleuilealaleuVal1Ala1aG1yleuA1aSerA1a 18
Db 17275 GCCTTGCTCATTCGATTGTCGCGGCTCTCGCTCGGCT 17313

RESULT 25
AAAB1490
ID AAAB1490 standard; DNA; 1437668 BP.
XX
AC AAAB1490;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
KM Meningococcus B; Memb; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
XX (CHIR ) CHIRON CORP.
PA
PI Frazer CM, Hickey E, Peterson J, Tetteijn H, Venter JC;
PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizsa M;
XX
XX WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
XX used in the diagnosis and treatment of N. meningitidis infection and
XX other Neisserial infections, for example, N.gonorrhoea -
XX
XX Claim 7; Page 866-1272; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
XX proteins from Neisseria genomic sequences. AAAB1453 to AAAB2414
XX represent specifically claimed Neisseria meningitidis genomic DNA
XX sequences; AAAB1260 to AAAB1303 and AAB25620 to AAB25663 represent
XX Neisseria DNA sequences and their corresponding proteins; AAAB1254 to
XX AAAB1259 and AAAB1304 to AAAB1321 represent PCR primers used in the
XX isolation of Neisseria meningitidis DNA sequences; and AAAB1322 to
XX AAAB1452 represent Neisseria meningitidis Memb polynucleotide ORF
XX sequences, which are all used in the exemplification of the present
XX invention. The nucleic acid sequences, protein sequences, and antibodies
XX against them, can be used in the manufacture of a composition. The
XX composition can be used as a medicament (or in the manufacture of a
XX medicament) for treating, preventing or diagnosing infection due to
XX Neisserial bacteria. For example, some of the identified proteins could
XX be components of vaccines against Meningococcus B; against all serotypes;
XX and/or against all pathogenic Neisseriae. Identification of sequences
XX from the bacterium will also facilitate production of biological probes,
XX particularly organism-specific probes. Attempts to make efficacious

```

CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX
 SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;
 Alignment Scores:
 Pred. No.: 1,21e+06 Length: 1437668
 Score: 43.00 Matches: 9
 Percent Similarity: 92.31% Conservative: 3
 Best Local Similarity: 69.23% Mismatches: 1
 Query Match: 52.44% Indels: 0
 DB: 21 Gaps: 0
 US-10-081-935-2 (1-18) x AAH53816 (1-1437668)
 QY 6 AAlaLeuIleAlaLeuValAlaIaGlyLeuAlaSerAla 18
 Db 1187664 GCCTTCTCATTCGATTCGCGGTCTCGCGCT 1187702
 RESULT 26
 AAH53816/C
 ID AAH53816 standard; DNA; 498 BP.
 XX
 AC AAH53816;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:3025.
 XX
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KM vaccination; endocarditis; ds.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN MO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAXO) GLAXO GROUP LTD.
 XX
 PI Kimmery MJ;
 DR WPI; 2001-316495/33.
 DR P-PSDB; AAG82966.
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX
 PS Claim 8; Page 796-797; 2188pp; English.
 XX
 CC AAH52304 to AAH53970 represent nucleic acids (i) encoding polypeptides
 CC (ii), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (i) and (ii) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (i) may be used to produce the
 CC S. epidermidis polypeptides (ii) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (ii) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used

CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 498 BP; 187 A; 60 C; 112 G; 139 T; 0 other;
 Alignment Scores:
 Pred. No.: 125 Length: 498
 Score: 42.00 Matches: 10
 Percent Similarity: 82.35% Conservative: 4
 Best Local Similarity: 58.82% Mismatches: 1
 Query Match: 51.22% Indels: 2
 DB: 22 Gaps: 1
 US-10-081-935-2 (1-18) x AAH53816 (1-498)
 QY 2 GlnPheSerHisAlaLeuIleAlaIaGlyLeuAlaSerAla 18
 Db 128 CAATTTTCACATTCCTTCTACGCTCAATCGCA-----TTGCGCTCATCT 84
 RESULT 27
 ABN91496/C
 ID ABN91496 standard; DNA; 504 BP.
 XX
 AC ABN91496;
 XX
 DT 24-JUL-2002 (first entry)
 XX
 DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:959.
 XX
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KM antibacterial; gene therapy; gene; ds.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN US6380370-B1.
 XX
 PD 30-APR-2002.
 XX
 PF 13-AUG-1998; 98US-0134001.
 XX
 PR 14-AUG-1997; 97US-055779P.
 PR 08-NOV-1997; 97US-064964P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Doucette-Stamm LA, Bush D;
 DR WPI; 2002-381255/41.
 DR P-PSDB; ABP38951.
 PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 XX
 PS Disclosure; SEQ ID 959; 267pp; English.
 XX
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.
 XX
 SQ Sequence 504 BP; 188 A; 61 C; 114 G; 141 T; 0 other;

Alignment Scores:

Pred. No.:	126	Length:	504
Score:	42.00	Matches:	10
Percent Similarity:	82.35%	Conservative:	4
Best Local Similarity:	58.82%	Mismatches:	1
Query Match:	51.22%	Indels:	2
DB:	24	Gaps:	1

US-10-081-935-2 (1-18) x ABN91496 (1-504)

OY 2 GlnpEsErHIsAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 134 CAATTTTCACATTTCTCTCTCTACGCTTAATGCA-----TTGGCCTCATCT 90

RESULT 28
AAS85360/C
ID AAS85360 standard; cDNA; 806 BP.
XX AAS85360;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE DNA encoding novel human diagnostic protein #21164.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX P1 Drmanac RT, Liu C, Tang YF;
XX
XX DR WPI; 2001-639362/73.
XX DR P-P8DB; ABG21173.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX
XX PS Claim 1; SEQ ID NO 21164; 103pp: English.
XX
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (II) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX	Sequence	806 BP; 187 A; 229 C; 213 G; 177 T; 0 other;
XX	Alignment Scores:	
XX	Pred. No.:	225
XX	Score:	42.00
XX	Percent Similarity:	78.57%
XX	Best Local Similarity:	64.29%
XX	Query Match:	51.22%
XX	DB:	23
XX	US-10-081-935-2 (1-18) x AAS85360 (1-806)	
XX	QY	3 PheSerHisAlaIleuLeuIleAlaIleuValAlaAlaIleuAla 16
XX	DB	317 TTCACCCACAGCTGCGACGCTCTGTCGACAGCGCTTTCGA 276
XX	RESULT 29	
XX	ID	ABQ79615/c
XX	AC	ABQ79615 standard; DNA; 881 BP.
XX	DT	25-NOV-2002 (first entry)
XX	DE	T. reesei cbh1 gene fragment.
XX	KM	Secreted protein; transcription; secretion stress; fungal; cbh1; promoter; ds.
XX	OS	Trichoderma reesei.
XX	PN	WO200264624-A2.
XX	PD	22-AUG-2002.
XX	PF	13-FEB-2002; 2002WO-FI00116.
XX	PR	13-FEB-2001; 2001FI-0000272.
XX	PA	(VALM) VALTTON TEKNIILINEN TUTKIMUSKESKUS.
XX	PI	Pakula T, Saloheimo M, Uusitalo J, Huuskonen A, Watson A;
XX	PI	Jeeones D, Archer D, Penttilae M;
XX	DR	WPI; 2002-657583/70.
XX	PT	New DNA sequence located in a promoter of a secreted protein, useful
XX	PT	for optimized protein production and design better fungal strains for
XX	PT	protein production
XX	PS	Disclosure; Page 84; 84pp; English.
XX	CC	The invention relates to a DNA sequence located in a promoter of a
XX	CC	secreted protein, where the DNA sequence mediates transcriptional down-
XX	CC	regulation of secreted proteins under secretion stress. The DNA sequence, a
XX	CC	promoter or the fungal host is useful to optimize protein production. The
XX	CC	methods are useful to or to modify fungal strains for other purposes and
XX	CC	selectively regulate the expression of certain undesired or desired
XX	CC	proteins in the host. The present sequence represents a T. reesei DNA
XX	CC	fragment located upstream of the cbh1 gene promoter. The promoters of the
XX	CC	invention are modified such that they lack the present sequence.
XX	Sequence	881 BP; 256 A; 193 C; 218 G; 214 T; 0 other;
XX	Alignment Scores:	
XX	Pred. No.:	250
XX	Score:	42.00
XX	Percent Similarity:	84.62%
XX	Best Local Similarity:	51.85%
XX	Query Match:	51.22%
XX	DB:	24
XX	Length:	881
XX	Matches:	7
XX	Conservative:	4
XX	Mismatches:	2
XX	Indels:	0
XX	Gaps:	0


```

XX AC AB079607;
XX XX 25-NOV-2002 (first entry)
XX DT
XX DE T. reesei cbh1 gene fragment.
XX XX
XX KW Secreted protein; transcription; secretion stress; fungal; cbh1;
XX KW promoter; ds.
XX OS Trichoderma reesei.
XX PN WO200264624-A2.
XX PD 22-AUG-2002.
XX PF 13-FEB-2002; 2002WO-FI00116.
XX PR 13-FEB-2001; 2001FI-0000272.
XX PA (VALM ) VALTION TEKNIILINEN TUTKIMUSKESKUS.
XX PI Pakula T, Saloheimo M, Uusitalo J, Huuskonen A, Watson A;
XX PI Jeenes D, Archer D, Penttilae M;
XX DR WPI; 2002-657583/70.
XX DR
XX PT New DNA sequence located in a promoter of a secretable protein, useful
XX PT for optimized protein production and design better fungal strains for
XX PT protein production
XX PS
XX SQ Disclosure; Page 81; 84pp; English.
XX CC The invention relates to a DNA sequence located in a promoter of a
XX CC secretable protein, where the DNA sequence mediates transcriptional down-
XX CC regulation of secreted proteins under secretion stress. The DNA sequence,
XX CC promoter or the fungal host is useful to optimize protein production. The
XX CC methods are useful to or to modify fungal strains for other purposes and
XX CC selectively regulate the expression of certain undesired or desired
XX CC proteins in the host. The present sequence represents a T. reesei DNA
XX CC fragment located upstream of the cbh1 gene promoter.
XX SQ Sequence 1201 BP; 328 A; 273 C; 301 G; 299 T; 0 other;

Alignment Scores:
Pred. No.: 366 Length: 1201
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
DB: 24 Gaps: 0

US-10-081-935-2 (1-18) x AB079607 (1-1201)
Qy 3 PheserHisAlaLeuIleAlaLeuValAlaIaGlyLeu 15
Db 452 TTTAGCCACTTGATTGTACATTATTACCGCTGTATA 414

RESULT 33
AB079605/c
ID AB079605 standard; DNA; 1281 BP.
XX AC AB079605;
XX XX
XX DT 25-NOV-2002 (first entry)
XX XX
XX DE T. reesei cbh1 gene fragment.
XX XX
XX KW Secreted protein; transcription; secretion stress; fungal; cbh1;
XX KW promoter; ds.
XX OS Trichoderma reesei.
XX PN
XX PD
XX PF
XX PR
XX PA (VALM ) VALTION TEKNIILINEN TUTKIMUSKESKUS.
XX PI Pakula T, Saloheimo M, Uusitalo J, Huuskonen A, Watson A;
XX PI Jeenes D, Archer D, Penttilae M;
XX DR WPI; 2002-657583/70.
XX DR
XX PT New DNA sequence located in a promoter of a secretable protein, useful
XX PT for optimized protein production and design better fungal strains for
XX PT protein production
XX PS
XX SQ Disclosure; Page 80; 84pp; English.
XX CC The invention relates to a DNA sequence located in a promoter of a
XX CC secretable protein, where the DNA sequence mediates transcriptional down-
XX CC regulation of secreted proteins under secretion stress. The DNA sequence,
XX CC promoter or the fungal host is useful to optimize protein production. The
XX CC methods are useful to or to modify fungal strains for other purposes and
XX CC selectively regulate the expression of certain undesired or desired
XX CC proteins in the host. The present sequence represents a T. reesei DNA
XX CC fragment located upstream of the cbh1 gene promoter.
XX SQ Sequence 1281 BP; 356 A; 290 C; 317 G; 318 T; 0 other;

Alignment Scores:
Pred. No.: 396 Length: 1281
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
DB: 24 Gaps: 0

US-10-081-935-2 (1-18) x AB079605 (1-1281)
Qy 3 PheserHisAlaLeuIleAlaLeuValAlaIaGlyLeu 15
Db 532 TTTAGCCACTTGATTGTACATTATTACCGCTGTATA 494

RESULT 34
AB079602/c
ID AB079602 standard; DNA; 1474 BP.
XX AC AB079602;
XX XX
XX DT 25-NOV-2002 (first entry)
XX XX
XX DE T. reesei cbh1 gene fragment.
XX XX
XX KW Secreted protein; transcription; secretion stress; fungal; cbh1;
XX KW promoter; ds.
XX OS Trichoderma reesei.
XX PN WO200264624-A2.
XX PD 22-AUG-2002.
XX PF 13-FEB-2002; 2002WO-FI00116.
XX PR 13-FEB-2001; 2001FI-0000272.
XX PA (VALM ) VALTION TEKNIILINEN TUTKIMUSKESKUS.
XX PI Pakula T, Saloheimo M, Uusitalo J, Huuskonen A, Watson A;
XX PI Jeenes D, Archer D, Penttilae M;

```

```

PN WO200264624-A2.
XX XX 22-AUG-2002.
XX PD
XX PF 13-FEB-2002; 2002WO-FI00116.
XX PR 13-FEB-2001; 2001FI-0000272.
XX PA (VALM ) VALTION TEKNIILINEN TUTKIMUSKESKUS.
XX PI Pakula T, Saloheimo M, Uusitalo J, Huuskonen A, Watson A;
XX PI Jeenes D, Archer D, Penttilae M;
XX DR WPI; 2002-657583/70.
XX DR
XX PT New DNA sequence located in a promoter of a secretable protein, useful
XX PT for optimized protein production and design better fungal strains for
XX PT protein production
XX PS
XX SQ Disclosure; Page 80; 84pp; English.
XX CC The invention relates to a DNA sequence located in a promoter of a
XX CC secretable protein, where the DNA sequence mediates transcriptional down-
XX CC regulation of secreted proteins under secretion stress. The DNA sequence,
XX CC promoter or the fungal host is useful to optimize protein production. The
XX CC methods are useful to or to modify fungal strains for other purposes and
XX CC selectively regulate the expression of certain undesired or desired
XX CC proteins in the host. The present sequence represents a T. reesei DNA
XX CC fragment located upstream of the cbh1 gene promoter.
XX SQ Sequence 1281 BP; 356 A; 290 C; 317 G; 318 T; 0 other;

Alignment Scores:
Pred. No.: 396 Length: 1281
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
DB: 24 Gaps: 0

US-10-081-935-2 (1-18) x AB079605 (1-1281)
Qy 3 PheserHisAlaLeuIleAlaLeuValAlaIaGlyLeu 15
Db 532 TTTAGCCACTTGATTGTACATTATTACCGCTGTATA 494

RESULT 34
AB079602/c
ID AB079602 standard; DNA; 1474 BP.
XX AC AB079602;
XX XX
XX DT 25-NOV-2002 (first entry)
XX XX
XX DE T. reesei cbh1 gene fragment.
XX XX
XX KW Secreted protein; transcription; secretion stress; fungal; cbh1;
XX KW promoter; ds.
XX OS Trichoderma reesei.
XX PN WO200264624-A2.
XX PD 22-AUG-2002.
XX PF 13-FEB-2002; 2002WO-FI00116.
XX PR 13-FEB-2001; 2001FI-0000272.
XX PA (VALM ) VALTION TEKNIILINEN TUTKIMUSKESKUS.
XX PI Pakula T, Saloheimo M, Uusitalo J, Huuskonen A, Watson A;
XX PI Jeenes D, Archer D, Penttilae M;

```

XX DR MPI: 2002-657583/70.
XX PT New DNA sequence located in a promoter of a secretable protein, useful
PT for optimized protein production and design better fungal strains for
PT protein production -
XX PS Disclosure; Page 79; 84pp; English.
XX CC The invention relates to a DNA sequence located in a promoter of a
CC secretable protein, where the DNA sequence mediates transcriptional down-
CC regulation of secreted proteins under secretion stress. The DNA sequence,
CC promoter or the fungal host is useful to optimize protein production. The
CC methods are useful to or to modify fungal strains for other purposes and
CC selectively regulate the expression of certain undesired or desired
CC proteins in the host. The present sequence represents a T. reesei DNA
CC fragment located upstream of the cbh1 gene promoter.
XX SQ Sequence 1474 BP; 397 A; 337 C; 357 G; 383 T; 0 other;

Alignment Scores:
Pred. No.: 470 Length: 1474
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
DB: 24 Gaps: 0

US-10-081-935-2 (1-18) x ABQ79602 (1-1474)

QY 3 PhseRh1aLeu1leAlaLeuVa1AlaAlaGlyLeu 15
Db 1466 TTTAGCCACTTGATGTGACATTATTAGCCCGCTGTATA 1428
RESULT 35
ABL02695
ID ABL02695 standard; cDNA; 1569 BP.
XX AC ABL02695;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2567.
XX KM Drosophila: developmental biology; cell signalling; insecticide;
XX KM pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR MPI: 2001-656860/75.
XX DR P-PSDB; ABB58592.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX PS Claim 1; SEQ ID NO 2567; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1569 BP; 399 A; 465 C; 381 G; 324 T; 0 other;

Alignment Scores:
Pred. No.: 507 Length: 1569
Score: 42.00 Matches: 8
Percent Similarity: 91.67% Conservative: 3
Best Local Similarity: 66.67% Mismatches: 1
Query Match: 51.22% Indels: 0
DB: 23 Gaps: 0

US-10-081-935-2 (1-18) x ABL02695 (1-1569)

QY 4 SerH1aLeu1leAlaLeuVa1AlaAlaGlyLeu 15
Db 807 TCCCACTCCGCTGTTCCCTCGTCCGACACTT 842
RESULT 36
ABQ79601/c
ID ABQ79601 standard; DNA; 1714 BP.
XX AC ABQ79601;
XX DT 25-NOV-2002 (first entry)
XX DE T. reesei cbh1 gene fragment.
XX KM Secreted protein; transcription; secretion stress; fungal; cbh1;
XX KM promoter; ds.
XX OS Trichoderma reesei.
XX PN WO200264624-A2.
XX PD 22-AUG-2002.
XX PF 13-FEB-2002; 2002WO-FI00116.
XX PR 13-FEB-2001; 2001FI-0000272.
XX PA (VALM) VALTION TEKNIILLINEN TUTKIMUSKESKUS.
XX PI Pakula T, Saloheimo M, Unsitalo J, Huuskonen A, Watson A;
XX PI Jeenes D, Archer D, Penttilae M;
XX DR MPI: 2002-657583/70.
XX PT New DNA sequence located in a promoter of a secretable protein, useful
PT for optimized protein production and design better fungal strains for
PT protein production -
XX PS Disclosure; Page 78; 84pp; English.
XX CC The invention relates to a DNA sequence located in a promoter of a
CC secretable protein, where the DNA sequence mediates transcriptional down-
CC regulation of secreted proteins under secretion stress. The DNA sequence,
CC promoter or the fungal host is useful to optimize protein production. The
CC methods are useful to or to modify fungal strains for other purposes and
CC selectively regulate the expression of certain undesired or desired
CC proteins in the host. The present sequence represents a T. reesei DNA
CC fragment located upstream of the cbh1 gene promoter.
XX SQ Sequence 1714 BP; 459 A; 389 C; 422 G; 444 T; 0 other;

Alignment Scores:

Pred. No.:	565	Length:	1714
Score:	42.00	Matches:	7
Percent Similarity:	84.62%	Conservative:	4
Best Local Similarity:	53.85%	Mismatches:	2
Query Match:	51.22%	Indels:	0
DB:	24	Gaps:	0

US-10-081-935-2 (1-18) x AB079601 (1-1714)

OY 3 PhseSerHisAlaLeuIleAlaValAlaIaGlyLeu 15
 Db 1466 TTTAGCCACTGATTGTAACATTATTAGCCGCTGGTATA 1428

RESULT 37
 AA058017/c
 ID AA058017 standard; DNA; 1776 BP.

AC AA058017;
 XX
 DT 25-MAR-2003 (updated)
 DT 14-SEP-1994 (first entry)
 XX
 DE Sequence of altered cellobiohydrolase 1 (cbh1) promoter of pMI-27.
 XX
 KW Promoter; cellobiohydrolase 1; cbh1; pMI-27; ss.
 XX
 OS Synthetic.
 XX

Key Location/Qualifiers
 FT misc_difference 704..709
 FT /*tag= a
 FT /label= altered sequence
 FT 716..748
 FT /*tag= b
 FT /label= polylinker
 FT 748..749
 FT misc_feature
 FT /*tag= C
 FT /label= deletion site
 FT misc_difference 1054..1059
 FT /*tag= d
 FT /label= altered sequence
 FT 1774..1776
 FT /*tag= e
 FT /label= start codon

MO9404673-A1.
 PN
 PD 03-MAR-1994.
 XX
 PF 19-AUG-1993; 93MO-F100330.
 XX
 PR 19-AUG-1992; 92US-0932485.
 XX
 PA (ALKO-) ALKO OY AB.
 XX
 PI Ilmen MH, Nakari TH, Nevalainen KM, Onnela M, Penttilä ME;
 XX
 DR WPI; 1994-083192/10.
 XX
 PT Cloning promoters active in a desired environmental condition -
 PT used partic. for expression of genes in Trichoderma fungal hosts
 PT in glucose-contg. medium
 XX
 PS Example; Figure 18B; 120pp; English.
 XX
 CC PM1016del10(2) was used as a PCR template to yield a 280 bp fragment
 CC E (primers = AA058023 and AA058024), beginning at the promoter
 CC internal polylinker and ending at -720, and a 720 bp fragment F
 CC (primers = AA058026 and AA058028), beginning at -720 and ending at
 CC KspI at -16. Fragments D and E were purified from agarose gel
 CC and digested with BstEII-XbaI and XbaI-KspI respectively, & ligated
 CC to the 7.8 kb BstEII-KspI fragment of pMI-25 to produce produce

CC pMI-27. The cbh1 promoter of pMI-27 has sequence alterations at
 CC posns. -1505-1500 (genomic sequence GTGGCG, altered sequence
 CC TCTAAA) and at posn. -720 to -715 (genomic sequence GTGGCG, altered
 CC sequence TCTNGA) upstream of the translation initiation codon of intact
 CC cbh1 promoter. The sequence of the altered cbh1 promoter in pMI-27
 CC is provided in AA058017.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX

Sequence 1776 BP; 490 A; 399 C; 430 G; 457 T; 0 other.

Alignment Scores:

Pred. No.:	590	Length:	1776
Score:	42.00	Matches:	7
Percent Similarity:	84.62%	Conservative:	4
Best Local Similarity:	53.85%	Mismatches:	2
Query Match:	51.22%	Indels:	0
DB:	15	Gaps:	0

US-10-081-935-2 (1-18) x AA058017 (1-1776)

OY 3 PhseSerHisAlaLeuIleAlaValAlaIaGlyLeu 15
 Db 1029 TTTAGCCACTGATTGTAACATTATTAGCCGCTGGTATA 991

RESULT 38
 AA058018/c
 ID AA058018 standard; DNA; 1776 BP.

AC AA058018;
 XX
 DT 25-MAR-2003 (updated)
 DT 14-SEP-1994 (first entry)
 XX
 DE Sequence of altered cellobiohydrolase 1 (cbh1) promoter of pMI-28.
 XX
 KW Promoter; cellobiohydrolase 1; cbh1; pMI-28; ss.
 XX
 OS Synthetic.
 XX

Key Location/Qualifiers
 FT misc_difference 704..709
 FT /*tag= a
 FT /label= altered sequence
 FT 716..748
 FT /*tag= b
 FT /label= polylinker
 FT 748..749
 FT misc_feature
 FT /*tag= C
 FT /label= deletion site
 FT misc_difference 773..778
 FT /*tag= d
 FT /label= altered sequence
 FT misc_difference 1054..1059
 FT /*tag= e
 FT /label= altered sequence
 FT 1774..1776
 FT /*tag= f
 FT /label= start codon

MO9404673-A1.
 PN
 PD 03-MAR-1994.
 XX
 PF 19-AUG-1993; 93MO-F100330.
 XX
 PR 19-AUG-1992; 92US-0932485.
 XX
 PA (ALKO-) ALKO OY AB.
 XX
 PI Ilmen MH, Nakari TH, Nevalainen KM, Onnela M, Penttilä ME;
 XX
 DR WPI; 1994-083192/10.
 XX

PT Cloning promoters active in a desired environmental condition -
PT used partic. for expression of genes in Trichoderma fungal hosts
PT in glucose-contg. medium
XX
PS Example; Figure 18C; 120pp; English.
XX
CC pMOIde10(2) was used as a PCR template to yield a 280 bp fragment
CC G (primers = AA058023 and AA058025), beginning at the promoter
CC internal polylinker and ending at -720, and a 720 bp fragment H
CC (primers = AA058026 and AA058028), beginning at -720 and ending at
CC Kapi at -16. Fragments G and H were purified from agarose gel
CC and digested with BstEII-XbaI and XbaI-KspI respectively, & ligated
CC to the 7.8 kb BstEII-KspI fragment of pMI-25 to produce produce
CC pMI-28. The cbh1 promoter of pMI-28 has sequence alterations at
CC posns. -1505-1500 (genomic sequence GTGGCG, altered sequence
CC TCTAAA), -1001-996 (genomic sequence GTGGCG, altered sequence
CC TCTAAA), and at posn. -720 to -715 (genomic sequence GTGGCG, altered
CC sequence TCTAGA) upstream of the translation initiation codon of intact
CC cbh1 promoter. The sequence of the altered cbh1 promoter in pMI-28
CC is provided in AA058018.
CC (updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 1776 BP; 490 A; 400 C; 429 G; 457 T; 0 other;
SO
XX
Alignment Scores:
Pred. No.: 590 Length: 1776
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
DB: 15 Gaps: 0
US-10-081-935-2 (1-18) x AA058018 (1-1776)
Qy 3 PhseSerHsAlaLeuIleAlaLeuValAlaAlaGlyLeu 15
Db 1029 TTTAGCCACTTGATTGATCAATTATTAAGCCCTGCTATA 991
RESULT 39
AB079594/C
ID AB079594 standard; DNA; 1824 BP.
XX
AC AB079594;
XX
XX 25-NOV-2002 (first entry)
XX
DE T. reesei cbh1 gene fragment.
XX
KW Secreted protein; transcription; secretion stress; fungal; cbh1;
KW promoter; ds.
XX
OS Trichoderma reesei.
XX
OS WO200264624-A2.
XX
XX
PD 22-AUG-2002.
XX
PF 13-FEB-2002; 2002WO-FI00116.
XX
PR 13-FEB-2001; 2001FI-0000272.
XX
XX (VALW) VALTION TEKNIILLINEN TUTKIMUSKESKUS.
XX
XX Pakula T, Saloheimo M, Uusitalo J, Huuskonen A, Watson A;
XX Pi Jeenes D, Archer D, Penttilae M;
XX
XX WPI: 2002-657583/70.
XX
XX New DNA sequence located in a promoter of a secretable protein, useful
XX PT for optimized protein production and design better fungal strains for
XX PT protein production -
XX
XX PS Disclosure; Page 74; 84pp; English.

XX
CC The invention relates to a DNA sequence located in a promoter of a
CC secretable protein, where the DNA sequence mediates transcriptional down-
CC regulation of secreted proteins under secretion stress. The DNA sequence,
CC promoter or the fungal host is useful to optimize protein production. The
CC methods are useful to or to modify fungal strains for other purposes and
CC selectively regulate the expression of certain undesired or desired
CC proteins in the host. The present sequence represents a T. reesei DNA
CC fragment located upstream of the cbh1 gene promoter.
XX
SQ Sequence 1824 BP; 492 A; 406 C; 452 G; 474 T; 0 other;
SO
XX
Alignment Scores:
Pred. No.: 610 Length: 1824
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
DB: 24 Gaps: 0
US-10-081-935-2 (1-18) x AB079594 (1-1824)
Qy 3 PhseSerHsAlaLeuIleAlaLeuValAlaAlaGlyLeu 15
Db 1466 TTTAGCCACTTGATTGATCAATTATTAAGCCCTGCTATA 1428
RESULT 40
AB079597/C
ID AB079597 standard; DNA; 1874 BP.
XX
AC AB079597;
XX
XX 25-NOV-2002 (first entry)
XX
DE T. reesei cbh1 gene fragment.
XX
KW Secreted protein; transcription; secretion stress; fungal; cbh1;
KW promoter; ds.
XX
OS Trichoderma reesei.
XX
OS WO200264624-A2.
XX
XX
PD 22-AUG-2002.
XX
PF 13-FEB-2002; 2002WO-FI00116.
XX
PR 13-FEB-2001; 2001FI-0000272.
XX
XX (VALW) VALTION TEKNIILLINEN TUTKIMUSKESKUS.
XX
XX Pakula T, Saloheimo M, Uusitalo J, Huuskonen A, Watson A;
XX Pi Jeenes D, Archer D, Penttilae M;
XX
XX WPI: 2002-657583/70.
XX
XX New DNA sequence located in a promoter of a secretable protein, useful
XX PT for optimized protein production and design better fungal strains for
XX PT protein production -
XX
XX PS Disclosure; Page 76; 84pp; English.
XX
XX The invention relates to a DNA sequence located in a promoter of a
XX secretable protein, where the DNA sequence mediates transcriptional down-
XX regulation of secreted proteins under secretion stress. The DNA sequence,
XX promoter or the fungal host is useful to optimize protein production. The
XX methods are useful to or to modify fungal strains for other purposes and
XX selectively regulate the expression of certain undesired or desired
XX proteins in the host. The present sequence represents a T. reesei DNA
XX fragment located upstream of the cbh1 gene promoter.
XX
XX
XX Sequence 1874 BP; 499 A; 420 C; 467 G; 488 T; 0 other;
XX
XX

Alignment Scores:

Pred. No.:	630	Length:	1874
Score:	42.00	Matches:	7
Percent Similarity:	84.62%	Conservative:	4
Best Local Similarity:	53.85%	Mismatches:	2
Query Match:	51.22%	Indels:	0
DB:	24	Gaps:	0

US-10-081-935-2 (1-18) x ABQ79597 (1-1874)

QY	3	PhSeRtHsAlaLeuIleAlaLeuValAlaAlaGlyLeu	15
Db	1466	TTTAGCCACTTGATTGTACAATTAATTAGCCCGCTGTATA	1428

Search completed: November 13, 2003, 12:08:30
 Job time : 842 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 13, 2003, 11:51:07 ; Search time 52 Seconds
(without alignments)

152.786 Million cell updates/sec

Title: US-10-081-935-2

Sequence: 1 MQFSHALIALVAAGLASA 18

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10081935/runat_12112003_144407_3656/app_query.fasta_1.199
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-NO=US10081935 @CGN_1_1_44 @runat_12112003_144407_3656 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	57.3	4403765	3	US-09-103-840A-2
2	44	57.3	4411529	3	US-09-103-840A-1
3	44	53.7	248	1	US-08-525-697-3
4	44	53.7	1302	1	US-08-525-697-1
5	44	53.7	10091	3	US-09-058-489-34
6	42	51.2	471	4	US-09-252-991A-8612
7	42	51.2	504	4	US-09-134-001C-959
8	42	51.2	1781	2	US-08-389-564B-21
9	42	51.2	1781	2	US-08-389-564B-22
10	42	51.2	1781	3	US-08-466-047B-21
11	42	51.2	1781	3	US-08-466-047B-22
12	42	51.2	2218	2	US-08-389-564B-18

C 13	42	51.2	2218	3	US-09-066-597-1	Sequence 1, Appli
C 14	42	51.2	2218	3	US-08-466-047B-18	Sequence 18, Appl
C 15	42	51.2	2266	2	US-08-389-564B-20	Sequence 20, Appl
C 16	42	51.2	2266	3	US-08-466-047B-20	Sequence 20, Appl
C 17	42	51.2	4176	3	US-09-254-733-1	Sequence 1, Appli
C 18	42	51.2	8075	3	US-08-374-077C-1	Sequence 1, Appli
C 19	42	51.2	8075	3	US-08-895-590-1	Sequence 1, Appli
C 20	42	51.2	8075	4	US-09-539-879A-1	Sequence 1, Appli
C 21	42	51.2	8096	3	US-09-058-489-33	Sequence 33, Appl
C 22	41	50.0	528	4	US-09-252-991A-9137	Sequence 9137, Ap
C 23	41	50.0	972	4	US-09-252-991A-15702	Sequence 15702, A
C 24	41	50.0	975	4	US-09-252-991A-15674	Sequence 15674, A
C 25	41	50.0	1056	4	US-09-252-991A-8728	Sequence 8728, Ap
C 26	41	50.0	1218	4	US-09-552-322-3	Sequence 3, Appli
C 27	41	50.0	1230	4	US-09-328-352-1918	Sequence 1918, Ap
C 28	41	50.0	1485	4	US-09-252-991A-15656	Sequence 15656, A
C 29	41	50.0	1566	4	US-09-252-991A-9220	Sequence 9220, Ap
C 30	41	50.0	1710	4	US-09-252-991A-9057	Sequence 9057, Ap
C 31	41	50.0	2123	4	US-09-724-864-18	Sequence 18, Appl
C 32	41	50.0	5045	3	US-09-390-721-1	Sequence 1, Appli
C 33	41	50.0	5045	3	US-09-390-721-3	Sequence 3, Appli
C 34	41	50.0	36412	4	US-08-311-731A-132	Sequence 132, App
C 35	40	48.8	304	4	US-08-313-294A-2891	Sequence 2891, Ap
C 36	40	48.8	465	4	US-09-252-991A-16275	Sequence 16275, A
C 37	40	48.8	504	4	US-09-512-342-19	Sequence 19, Appl
C 38	40	48.8	669	4	US-09-252-991A-3703	Sequence 3703, Ap
C 39	40	48.8	717	3	US-08-913-014A-7	Sequence 7, Appli
C 40	40	48.8	765	4	US-09-252-991A-9931	Sequence 9931, Ap
C 41	40	48.8	795	3	US-08-913-014A-8	Sequence 8, Appli
C 42	40	48.8	951	4	US-09-252-991A-16386	Sequence 16386, A
C 43	40	48.8	1053	4	US-09-252-991A-6421	Sequence 6421, Ap
C 44	40	48.8	1092	4	US-09-252-991A-16212	Sequence 16212, A
C 45	40	48.8	1179	4	US-09-252-991A-10081	Sequence 10081, A

ALIGNMENTS

RESULT 1

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 1.95e+05
Score: 47.00
Percent Similarity: 92.86%
Best Local Similarity: 64.29%
Query Match: 57.32%
DB: 3
Length: 4403765
Matches: 9
Conservative: 4
Indels: 1
Gaps: 0

US-10-081-935-2 (1-18) x US-09-103-840A-2 (1-4403765)

```
Oy 3 PheSerHisAlaLeuValAlaAlaGlyLeuAla 16
|||||
Db 1218600 TTCCGCACGCCCTGGTGCATCGTTGCCGTGAATTGCG 1218641
|||||

RESULT 2
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 1.85e+05 Length: 4411529
Score: 47.00 Matches: 9
Percent Similarity: 92.86% Conservative: 4
Best Local Similarity: 64.29% Mismatches: 1
Query Match: 57.32% Indels: 0
DB: 3 Gaps: 0

US-10-081-935-2 (1-18) x US-09-103-840A-1 (1-4411529)

Oy 3 PheSerHisAlaLeuValAlaAlaGlyLeuAla 16
|||||
Db 1219072 TTCCGCACGCCCTGGTGCATCGTTGCCGTGAATTGCG 1219113
|||||

RESULT 3
US-08-525-697-3
; Sequence 3, Application US/08525697
; Patent No. 5795764
; GENERAL INFORMATION:
; APPLICANT: Christgau, Stephan
; APPLICANT: Andersen, Lene N
; APPLICANT: Kauppinen, Sakari
; APPLICANT: Heldt-Hansen, Hans P
; APPLICANT: Dalboege, Henrik
; TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5795764o No. 5795764disk of No. 5795764th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,697
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4004.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0486/93
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1302 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0486/93
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-525-697-3

Alignment Scores:
Pred. No.: 3.04 Length: 248
Score: 44.00 Matches: 8
Percent Similarity: 77.78% Conservative: 6
Best Local Similarity: 44.44% Mismatches: 4
Query Match: 53.66% Indels: 0
DB: 1 Gaps: 0

US-10-081-935-2 (1-18) x US-08-525-697-3 (1-248)

Oy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18
|||||
Db 24 ATGAAGCTTTCTTCACATGCTCTCCAGCCTCGCCGCTGGGGTGGCGACGCT 77
|||||

RESULT 4
US-08-525-697-1
; Sequence 1, Application US/08525697
; Patent No. 5795764
; GENERAL INFORMATION:
; APPLICANT: Christgau, Stephan
; APPLICANT: Andersen, Lene N
; APPLICANT: Kauppinen, Sakari
; APPLICANT: Heldt-Hansen, Hans P
; APPLICANT: Dalboege, Henrik
; TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5795764o No. 5795764disk of No. 5795764th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,697
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4004.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0486/93
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1302 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

[illegible]

CDS
/product="virion glycoprotein RL10"
/protein_id="AA000650.1"
/db_xref="GI:19881030"
/translation="MRVYTLGSLTLLLAIGATAIHNDVCTSLDGTQLLCKNRNNW
SNPKPTATSGTHICTLANCTCKRYTEPLPLISVLGIYSAMGAGSPIATLIILLVIFV
IYSRERADDONGTVEDPFLAYRDLTRKKLDSHASKQNIVERIPYPCRLRDDEPP
LITNAEDDEEDNF"
7202..7900
/note="RL11 family"
/codon_start=1
/product="IgG Fc-binding glycoprotein RL11"
/protein_id="AA000651.1"
/db_xref="GI:19881031"
/translation="MKRDALIKYVVWSTAFCLISAKPTPTTDAIGLPRLRLDTYTCE
TRNRTHASRLNTHGSVSWRYALPGHKPTQTMPLCSITDAASHTRDRVKARFICTR
LSPATLYNVTVEDMGVILVDTEYDGAESFYFFIISRAVCEAYATCAFWKLDNRQCT
HKMAGTNPDPDLDMWMLNISTHCAWIAGLLIIIVASLVGLLRLRRVGVQNAVYRLSN
RDTEPLQDAYSE"
7939..9021
/note="RL11 family"
/codon_start=1
/product="glycoprotein RL12"
/protein_id="AA000652.1"
/db_xref="GI:19881032"
/translation="MSETCSVATLLVLLALSQLEEDNTNNSVNTSTNDSNLSFTT
TSSPKATSTPTSPFTNLTSTSTYLNVSTVNSTHTPTLTTLPTSTTSHSFQNTIIT
EYSTYNUCNETTEVSSNSTESITATSCNTANTITFDNTISTAIQTITTEIVIKAW
TVHKGESNTLVHNSVTFQNCRTVMNHPYNIITHKVRKENKPHLCEASSGSHSVKH
NHKLCLOCTPKNLTFLDMVHNSGYIAECDGQNHQAQGFILTVNATHITNTRVC
QPKDVTDPITSLFHPNYSYOGHYPCASANSRGMAGVGLFLMACVLLFFARRVY
KKKYRPLRDVSESEFVRYTPEHD"
9123..10124
/note="RL11 family"
/codon_start=1
/product="glycoprotein RL13"
/protein_id="AA000653.1"
/db_xref="GI:19881033"
/translation="MVDHFAITWTLVLTLAGTVTSSCHTSHSVSTTVATTSTPS
VNTNSADTSNTSTESST
TTTTTTTST
HTARWVNTKETDLCFLGPDYTSPOAGICEKHWQSMITIDVTTENAGNITVR
HGDGHHYDKGYRLQVTHHTGTNRKCPNDTSYTPDHDKNETKTIENEFGMNYD
QPTFPMGHAIWAAVVVALLALYMGSRSSVTIVKGGKPKYKLSNKDPDEYWASP"
complement(10256..10432)
/note="contains hydrophobic domain"
/codon_start=1
/product="UL2"
/protein_id="AA000654.1"
/db_xref="GI:19881034"
/translation="MRGDAVSILIVEDLDLPSFGSFNASHAYISFRVLRGIFYVTVVV
WVLMWIKLLRONFP"
10504..11097
/note="RL11 family; contains hydrophobic domain"
/codon_start=1
/product="UL4"
/protein_id="AA000655.1"
/db_xref="GI:19881035"
/translation="MRRRCWGRPGCGATFKYGLLVLTWSVIGIHYGAYDHHVDH
RVIKSGNDWITSPYHLGSPFWYKRNATGTPVCYSVKTITYPDVRVCEKHHMNI
TIFNTEGWIYSLVTESPFLNDNKHYTEAFYVITDRNNGTEBELDYDDYDGSF
GSLGLESSGLGSGSGNEETSGSGDEIQNGFLITY"
11469..11924
/note="RL11 family; contains hydrophobic domain"
/codon_start=1
/product="UL5"
/protein_id="AA000656.1"
/db_xref="GI:19881036"
/translation="MFLGYSYDFMRFGFALCRAPGHYQTSVTTLLTCAVSSSLTAS
TSAAGSTGGYRLNLTORANLSSPDPTTSTLPRQIAAPPAHSSGLFPEOFYGLAIHA
AWTAFVLVGLIFFANVFPQITHGYRLDRDHRTLFGRRYQRLNS"
12130..12981
/note="RL11 family"
/codon_start=1

CDS
/product="glycoprotein UL6"
/protein_id="AA000657.1"
/db_xref="GI:19881037"
/translation="MPPPSFGCGLNKVMICSCILLVAVCAPRGSPHWSRVRRT
SLSPSCSPVNCTGVGDSVNSTIPVACNSTEWGRYNSFWPQLCWGWHRLSGQ
NLRLSLCSROHLTLHSTSDGYGTGYVYGNNCPHHHEIMKTCFNLTLVAKPTTTAAP
TTTSVPPTLLTLTRNTVIVGSPANTVLSSTWTTTAGNASVANGMEYROYQPLSLA
QTHRTAPLNATDNAEDVANLVATYASWGLVLLLLATVLVFLDLGTPQTAWRWRDOGR
EEQHLLL"
13052..13693
/note="RL11 family"
/codon_start=1
/product="glycoprotein UL7"
/protein_id="AA000658.1"
/db_xref="GI:19881038"
/translation="MAADPHITLMSGLCNKALWLSICLLIVAFLLASSSTTTWYV
SGPRLPLSLHANFTQDDVHMYRLKNETVSLCSVMVGDVFIARSNNSLGLCPMQL
ILFMTVNDGGLYVNGTNGTTRTSTTVFNVTIGLOFAPKTTKPKTKTKRKLSE
RTGMSLFRYARDLDSVRRKDDDDNIHLGLVAGLVALVIVCLMGNLKLCTH"
13799..14305
/note="RL11 family"
/codon_start=1
/product="glycoprotein UL8"
/protein_id="AA000659.1"
/db_xref="GI:19881039"
/translation="MNVTGQPPSPPPPGPSESTWNTTGGEDGERSVTTTVAVST
RPTLSGKTASTATSTAPVTTTRTTRTTRTSTSTSTSTSTSTSTSTSTSTSTSTST
LRTLSYHTTVAQPKIAVHYTWLILFVLVILFCLRIPOKLYDKWNRNRYGQ
VYVDTEL"
14327..14860
/note="RL11 family"
/codon_start=1
/product="glycoprotein UL9"
/protein_id="AA000660.1"
/db_xref="GI:19881040"
/translation="MFTYFSLTYVCKNLHEVPRECKVKTGYIGONVSLKSIHVM
SNSTWSVHTVSSLYMCKERESNDAPFPYNTDIOFNCTEELLLELLELTQYTA
RGYDTPRPCLMISTYCNVTNRTYAEKHTFASSIWIPIPPVAVVTLIILLSVIKIPI
RIMEWQRYRDTVYT"
14959..15657
/note="RL11 family"
/codon_start=1
/product="glycoprotein UL10"
/protein_id="AA000661.1"
/db_xref="GI:19881041"
/translation="MPFGSYVNCGFLTWFLVSAVKYEFSEACEEHTTYETHPLL
GKPLLAGHQTAYWYKHKCDSSNDKLTPLCYPPRNSKGLDESAENTFINYLCSN
STLIIARLNLTDAGEYCKKNSTRDSWPTMTCYRVIVRPITKTTSKKPESGPTRVKR
TTYDNYSVTLLLEELLFQTHRSTQNAHIAWLFIAVVIIVILFFFKIPQKLPDKWL
HKSIAKTGHFSAPH"
15758..16468
/note="RL11 family"
/codon_start=1
/product="glycoprotein UL11"
/protein_id="AA000662.1"
/db_xref="GI:19881042"
/translation="MCHRRLEWFKTSFYAMICVFTVSTDLQNAACVBEKARQGSNVT
FDARIYSGKNOSTYFRRGSYNADMNFTGNLTVHSTEIRLNYTCQNFSLTLNVT
SEYKNKYTRTTRSTSTNSFDTCFLKTVTVIKPKPKRTTLTKRQEPVTKPKTT
YVDYRGDGAFTIAVSPGIVSSQRSQATTISIWICLSIVIVVILFVFPRIPOKLLP
FMCPRPKDTELMIPTTEL"
16623..18392
/codon_start=1
/product="UL13"
/protein_id="AA000663.1"
/db_xref="GI:19881043"
/translation="MVVSVRGLTSYLPLLLCKVPPILLTWPSAHGLDIDVEELSH
DYEISVVRFAWQVGMQLPAPIMGRSEIDEORPSSPAQADVVDVSRPPRPRP
RLQSPYHQHQPSPGIRWOHEELQFLQTRQELLQOREEREQRQHQRRRYRWR
RPTFPFPPSPYPARSSTLIPARFARPSLANGAEQLNARGYGERRASRGNASPEP
SVEEYQYRVVGVNTRRRHHPHQRSYRRNALVANGRDSLLARLIRHQRTFTGY
RCRYQYRNRGSRARRDDTGSAAGHGGRPGGAGFSOLRERIVTDLOLFLRCHGT
RQASRRIRTRWEENTVMSDAASRLRAWFSRRTTYWQRTWTPGENPSAEAGELAVPPA

ADGVKETSQSMVTTEGEERQKTEORVEKAAKGGSETIEKIVETDTRVAETKXVE
 EVQVVEENNVVEETREDRDDEDAWIIPWGEWMDLSSLETTAAETEEVAVEK
 EETTKGADATTVTASSETDDVIQTNELPCELNNAEISGRVAVGTCPREGGPHR
 SFFRLCLGLWASSHLARRAITSVS"
 18719. .19780

CDS

Alignment Scores:

Pred. No.: 942 Length: 241087
 Score: 52.00 Matches: 10
 Percent Similarity: 93.75% Conservative: 5
 Best Local Similarity: 62.50% Mismatches: 1
 Query Match: 63.41% Indels: 0
 DB: 14 Gaps: 0

US-10-081-935-2 (1-18) x AF480884 (1-241087)

Qy 2 GlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 17

Db 66475 CGTTACGACGCGCTGGTGTGCTGCTGCCCGCCGACGCGAC 66428

RESULT 6

CNS01AF9 660 bp mRNA linear PLN 02-SEP-1999
 LOCUS Botrytis cinerea strain T4 cDNA library under conditions of
 DEFINITION nitrogen deprivation.

ACCESSION AL112861

VERSION AL112861.1 GI:5827480

KEYWORDS cDNA library; nitrogen deprivation.

SOURCE Botryotinia fuckeliana

ORGANISM Botryotinia fuckeliana

REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;

AUTHORS Helotiales; Sclerotiniaceae; Botryotinia.

TITLE Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.

JOURNAL Direct Submission

REFERENCE Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,

AUTHORS 78026 Versailles, France

TITLE 2 (bases 1 to 660)

JOURNAL Genoscope.

COMMENT Direct Submission

Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
 CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

The cDNA library to be analyzed within the framework of this
 project was created using a Botrytis cinerea strain which was grown
 under conditions of nitrogen deprivation, which is the normal
 situation for B. cinerea during its development on its host plant.
 The library was produced in an oriented direction, in the pBSII
 vector.

FEATURES

source Location/Qualifiers
 1. .660
 /organism="Botryotinia fuckeliana"
 /mol_type="mRNA"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID : W05H031"

BASE COUNT 161 a 214 c 131 g 154 t

ORIGIN

Alignment Scores:

Pred. No.: 13.6 Length: 660
 Score: 48.00 Matches: 10
 Percent Similarity: 72.22% Conservative: 3
 Best Local Similarity: 55.56% Mismatches: 5
 Query Match: 58.54% Indels: 0
 DB: 8 Gaps: 0

US-10-081-935-2 (1-18) x CNS01AF9 (1-660)

Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18

Db 81 ATGCAATTCTCATCTGCCATCATCTCTGCCATCACCGTTGCTCTTGCATCGGCC 134

RESULT 7

CNS01AO7

LOCUS

DEFINITION

Botrytis cinerea strain T4 cDNA library under conditions of

nitrogen deprivation.

ACCESSION AL113183

VERSION AL113183.1 GI:5827802

KEYWORDS cDNA library; nitrogen deprivation.

SOURCE Botryotinia fuckeliana

ORGANISM Botryotinia fuckeliana

REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;

AUTHORS Helotiales; Sclerotiniaceae; Botryotinia.

TITLE Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.

JOURNAL Direct Submission

REFERENCE Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,

AUTHORS 78026 Versailles, France

TITLE 2 (bases 1 to 660)

JOURNAL Genoscope.

COMMENT Direct Submission

Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
 CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

The cDNA library to be analyzed within the framework of this
 project was created using a Botrytis cinerea strain which was grown
 under conditions of nitrogen deprivation, which is the normal
 situation for B. cinerea during its development on its host plant.
 The library was produced in an oriented direction, in the pBSII
 vector.

FEATURES

source Location/Qualifiers
 1. .660
 /organism="Botryotinia fuckeliana"
 /mol_type="mRNA"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID : W59D121"

BASE COUNT 160 a 215 c 132 g 153 t

ORIGIN

Alignment Scores:

Pred. No.: 13.6 Length: 660
 Score: 48.00 Matches: 10
 Percent Similarity: 72.22% Conservative: 3
 Best Local Similarity: 55.56% Mismatches: 5
 Query Match: 58.54% Indels: 0
 DB: 8 Gaps: 0

US-10-081-935-2 (1-18) x CNS01AO7 (1-660)

Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18

Db 79 ATGCAATTCTCATCTGCCATCATCTCTGCCATCACCGTTGCTCTTGCATCGGCC 132

RESULT 8

CNS01B92

LOCUS

DEFINITION

Botrytis cinerea strain T4 cDNA library under conditions of

nitrogen deprivation.

ACCESSION AL113934

VERSION AL113934.1 GI:5828553

KEYWORDS cDNA library; nitrogen deprivation.

SOURCE Botryotinia fuckeliana

ORGANISM Botryotinia fuckeliana

REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;

AUTHORS Helotiales; Sclerotiniaceae; Botryotinia.

TITLE Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.

JOURNAL Direct Submission

REFERENCE Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,

AUTHORS 78026 Versailles, France

TITLE 2 (bases 1 to 660)

JOURNAL Genoscope.

COMMENT Direct Submission

CNS01AO7 660 bp mRNA linear PLN 02-SEP-1999
 LOCUS Botrytis cinerea strain T4 cDNA library under conditions of
 DEFINITION nitrogen deprivation.

ACCESSION AL113183

VERSION AL113183.1 GI:5827802

KEYWORDS cDNA library; nitrogen deprivation.

SOURCE Botryotinia fuckeliana

ORGANISM Botryotinia fuckeliana

REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;

AUTHORS Helotiales; Sclerotiniaceae; Botryotinia.

TITLE Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.

JOURNAL Direct Submission

REFERENCE Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,

AUTHORS 78026 Versailles, France

TITLE 2 (bases 1 to 660)

JOURNAL Genoscope.

COMMENT Direct Submission

Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :
 CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

The cDNA library to be analyzed within the framework of this
 project was created using a Botrytis cinerea strain which was grown
 under conditions of nitrogen deprivation, which is the normal
 situation for B. cinerea during its development on its host plant.
 The library was produced in an oriented direction, in the pBSII
 vector.

source Location/Qualifiers
 1. .660
 /organism="Botryotinia fuckeliana"
 /mol_type="mRNA"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID : W59D121"

BASE COUNT 160 a 215 c 132 g 153 t

ORIGIN

Alignment Scores:

Pred. No.: 13.6 Length: 660
 Score: 48.00 Matches: 10
 Percent Similarity: 72.22% Conservative: 3
 Best Local Similarity: 55.56% Mismatches: 5
 Query Match: 58.54% Indels: 0
 DB: 8 Gaps: 0

US-10-081-935-2 (1-18) x CNS01AO7 (1-660)

Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18

Db 79 ATGCAATTCTCATCTGCCATCATCTCTGCCATCACCGTTGCTCTTGCATCGGCC 132

RESULT 8

CNS01B92

LOCUS

DEFINITION

Botrytis cinerea strain T4 cDNA library under conditions of

nitrogen deprivation.

ACCESSION AL113934

VERSION AL113934.1 GI:5828553

KEYWORDS cDNA library; nitrogen deprivation.

SOURCE Botryotinia fuckeliana

ORGANISM Botryotinia fuckeliana

REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;

AUTHORS Helotiales; Sclerotiniaceae; Botryotinia.

TITLE Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.

JOURNAL Direct Submission

REFERENCE Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,

AUTHORS 78026 Versailles, France

TITLE 2 (bases 1 to 660)

JOURNAL Genoscope.

COMMENT Direct Submission

JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage : CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT The cDNA library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.

FEATURES

source

1..660 Location/Qualifiers
 /organism="Botryotinia fuckeliana"
 /mol_type="mRNA"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID : W21F101"

BASE COUNT 159 a 203 c 144 g 154 t

ORIGIN

Alignment Scores:
 Pred. No.: 13.6 Length: 660
 Score: 48.00 Matches: 10
 Percent Similarity: 72.22% Conservative: 3
 Best Local Similarity: 55.56% Mismatches: 5
 Query Match: 58.54% Indels: 0
 DB: 8 Gaps: 0

US-10-081-935-2 (1-18) x CNS01B92 (1-660)

Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18

Db 39 ATGCAATTCTCATCTGCCATCATCTCTGCCATCACCCTTGCTTGTGCATCGGCC 92

RESULT 9

CNS01C2S
 LOCUS Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation.
 DEFINITION

ACCSSION AL115004.1 GI:5829623

VERSION cDNA library; nitrogen deprivation.

KEYWORDS Botryotinia fuckeliana

SOURCE Botryotinia fuckeliana

ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;

REFERENCE 1 (bases 1 to 660)

AUTHORS Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.

TITLE Direct Submission

JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,

REFERENCE 78026 Versailles, France

2 (bases 1 to 660)

Genoscope.

TITLE Direct Submission

JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :

CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT The cDNA library to be analyzed within the framework of this

project was created using a Botrytis cinerea strain which was grown

under conditions of nitrogen deprivation, which is the normal

situation for B. cinerea during its development on its host plant.

The library was produced in an oriented direction, in the pBSII

vector.

FEATURES

source

1..660 Location/Qualifiers
 /organism="Botryotinia fuckeliana"
 /mol_type="mRNA"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID : W57B021"

BASE COUNT 158 a 203 c 145 g 154 t

ORIGIN

Alignment Scores:

Pred. No.: 13.6 Length: 660
 Score: 48.00 Matches: 10
 Percent Similarity: 72.22% Conservative: 3
 Best Local Similarity: 55.56% Mismatches: 5
 Query Match: 58.54% Indels: 0
 DB: 8 Gaps: 0

US-10-081-935-2 (1-18) x CNS01C2S (1-660)

Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18

Db 38 ATGCAATTCTCATCTGCCATCATCTCTGCCATCACCCTTGCTTGTGCATCGGCC 91

RESULT 10

CNS01CQG

LOCUS

DEFINITION

Botrytis cinerea strain T4 cDNA library under conditions of

nitrogen deprivation.

ACCSSION AL115856.1 GI:5831072

VERSION cDNA library; nitrogen deprivation.

KEYWORDS Botryotinia fuckeliana

SOURCE Botryotinia fuckeliana

ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;

REFERENCE Helotiales; Sclerotiniaceae; Botryotinia.

1 (bases 1 to 696)

AUTHORS Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.

TITLE Direct Submission

JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,

REFERENCE 78026 Versailles, France

2 (bases 1 to 696)

Genoscope.

TITLE Direct Submission

JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :

CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT The cDNA library to be analyzed within the framework of this

project was created using a Botrytis cinerea strain which was grown

under conditions of nitrogen deprivation, which is the normal

situation for B. cinerea during its development on its host plant.

The library was produced in an oriented direction, in the pBSII

vector.

FEATURES

source

1..696 Location/Qualifiers
 /organism="Botryotinia fuckeliana"
 /mol_type="mRNA"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID : W32D051"

BASE COUNT 174 a 216 c 142 g 164 t

ORIGIN

Alignment Scores:

Pred. No.: 14.3 Length: 696
 Score: 48.00 Matches: 10
 Percent Similarity: 72.22% Conservative: 3
 Best Local Similarity: 55.56% Mismatches: 5
 Query Match: 58.54% Indels: 0
 DB: 8 Gaps: 0

US-10-081-935-2 (1-18) x CNS01CQG (1-696)

Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18

Db 80 ATGCAATTCTCATCTGCCATCATCTCTGCCATCACCCTTGCTTGTGCATCGGCC 133

RESULT 11

CNS01DHA

LOCUS

DEFINITION

Botrytis cinerea strain T4 cDNA library under conditions of

nitrogen deprivation.

ACCSSION AL116822

VERSION AL116822.1 GI:5832038

KEYWORDS cDNA library; nitrogen deprivation.
SOURCE Botryotinia fuckeliana
ORGANISM Botryotinia fuckeliana
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;
 Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE 1 (bases 1 to 696)
AUTHORS Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France
REFERENCE 2 (bases 1 to 696)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage : CP 5706 91057 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT The cDNA library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.

FEATURES Location/Qualifiers
 source
 1..696
 /organism="Botryotinia fuckeliana"
 /mol_type="mRNA"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID : W44G081"
BASE COUNT 175 a 212 c 148 g 161 t
ORIGIN
 Alignment Scores:
 Pred. No.: 14.3 Length: 696
 Score: 48.00 Matches: 10
 Percent Similarity: 72.22% Conservative: 3
 Best Local Similarity: 55.56% Mismatches: 5
 Query Match: 58.54% Indels: 0
 DB: 8 Gaps: 0
 US-10-081-935-2 (1-18) x CNS01DHA (1-696)

QY 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18
 |||||
Db 53 ATGCAATTCTCATCTGCCATCATCTCTGCCATCACCGTTGCTTGCATCGGCC 106
 |||||

RESULT 12
CNS01BOR
LOCUS CNS01BOR 697 bp mRNA linear PLN 02-SEP-1999
DEFINITION Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation.
ACCESSION AL114499
VERSION AL114499.1 GI:5829118
KEYWORDS cDNA library; nitrogen deprivation.
SOURCE Botryotinia fuckeliana
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;
 Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE 1 (bases 1 to 697)
AUTHORS Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France
REFERENCE 2 (bases 1 to 697)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage : CP 5706 91057 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT The cDNA library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant.

The library was produced in an oriented direction, in the pBSII vector.
FEATURES Location/Qualifiers
 source
 1..697
 /organism="Botryotinia fuckeliana"
 /mol_type="mRNA"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID : W15F041"
BASE COUNT 177 a 219 c 151 g 149 t
ORIGIN
 Alignment Scores:
 Pred. No.: 14.3 Length: 697
 Score: 48.00 Matches: 10
 Percent Similarity: 72.22% Conservative: 3
 Best Local Similarity: 55.56% Mismatches: 5
 Query Match: 58.54% Indels: 0
 DB: 8 Gaps: 0
 US-10-081-935-2 (1-18) x CNS01BOR (1-697)

QY 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18
 |||||
Db 59 ATGCAATTCTCATCTGCCATCATCTCTGCCATCACCGTTGCTTGCATCGGCC 112
 |||||

RESULT 13
CNS01AOB
LOCUS CNS01AOB 720 bp mRNA linear PLN 02-SEP-1999
DEFINITION Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation.
ACCESSION AL113187
VERSION AL113187.1 GI:5827806
KEYWORDS cDNA library; nitrogen deprivation.
SOURCE Botryotinia fuckeliana
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomyces;
 Helotiales; Sclerotiniaceae; Botryotinia.
REFERENCE 1 (bases 1 to 720)
AUTHORS Bitton,F., Levis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France
REFERENCE 2 (bases 1 to 720)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage : CP 5706 91057 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT The cDNA library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.

FEATURES Location/Qualifiers
 source
 1..720
 /organism="Botryotinia fuckeliana"
 /mol_type="mRNA"
 /strain="T4"
 /db_xref="taxon:40559"
 /note="Genoscope sequence ID : W59D122"
BASE COUNT 180 a 221 c 145 g 174 t
ORIGIN
 Alignment Scores:
 Pred. No.: 14.8 Length: 720
 Score: 48.00 Matches: 10
 Percent Similarity: 72.22% Conservative: 3
 Best Local Similarity: 55.56% Mismatches: 5
 Query Match: 58.54% Indels: 0
 DB: 8 Gaps: 0

US-10-081-935-2 (1-18) x CNS01A0B (1-720)

Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSerAla 18
 Db 79 ATGCAATTCTCATCTGCCATCATCTCTGCCATCATCAGCTTCTTGCATCGGCC 132
 RESULT 14
 AC010228 78680 bp DNA linear PRI 16-JUL-2002
 LOCUS Homo sapiens chromosome 5 clone CTC-298B17, complete sequence.
 AC010228
 AC010228.8 GI:21844555
 HTG: Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 78680)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 78680)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 3 (bases 1 to 78680)
 DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 4 (bases 1 to 78680)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (01-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 5 (bases 1 to 78680)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 On Jul 16, 2002 this sequence version replaced gi:21637453.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.bngc.stanford.edu
 Quality: Phrap Quality >=40 99.9% of Sequence;
 Estimated Total Number of Errors is 0.
 NOTE: This insert is not the entire sequence of the clone (entire
 sequence is 145.4kb). It is clipped at the overlap with AC008955.
 The number of bases overlapped is 9996.
 FEATURES
 source
 1. 78680
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTC-298B17"
 BASE COUNT 26432 a 14896 c 13749 g 23603 t
 ORIGIN
 Alignment Scores: Length: 78680
 Pred. No.: 1.74e+03 Matches: 9
 Score: 48.00
 Percent Similarity: 82.35% Conservative: 5
 Best Local Similarity: 52.94% Mismatches: 3
 Query Match: 58.54% Indels: 0
 DB: 9 Gaps: 0
 US-10-081-935-2 (1-18) x AC010228 (1-78680)
 Qy 1 MetGlnPheSerHisAlaLeuValAlaAlaGlyLeuAlaSer 17
 Db 79 ATGCAATTCTCATCTGCCATCATCTCTGCCATCATCAGCTTCTTGCATCGGCC 132

Db 5755 CTCCAATTTACAAGGATATTACTGACTTTAGTATCTGCAGGAATGCCAGT 5805
 RESULT 15
 AC130471 148008 bp DNA linear HTG 10-AUG-2002
 LOCUS Homo sapiens chromosome 5 clone RP11-420M3, WORKING DRAFT SEQUENCE.
 DEFINITION 3 unordered pieces.
 AC130471
 AC130471.1 GI:22203244
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 148008)
 DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 5
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 148008)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (10-AUG-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

 Project Information
 Center Project Name: 567936
 Center clone name: RPCI-11_420M3

 Summary Statistics
 Consensus quality: 146731 bases at least Q40
 Consensus quality: 146937 bases at least Q30
 Consensus quality: 147107 bases at least Q20
 Estimated insert size: 0; null estimation
 Estimated insert size: 147808; sum-of-contigs estimation
 Quality coverage: 2.1474836E7 in Q20 bases; null estimation
 Quality coverage: 30.25 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1000: contig of 1000 bp in length
 * 1001 1100: gap of unknown length
 * 1101 2353: contig of 1253 bp in length
 * 2354 2453: gap of unknown length
 * 2454 148008: contig of 145555 bp in length.
 FEATURES
 source
 1. 148008
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="RP11-420M3"
 /clone.lib="RPCI human BAC library 11"
 BASE COUNT 50050 a 27259 c 25106 g 45393 t 200 others
 ORIGIN
 Alignment Scores: Length: 148008
 Pred. No.: 3.31e+03 Matches: 9
 Score: 48.00
 Percent Similarity: 82.35% Conservative: 5
 Best Local Similarity: 52.94% Mismatches: 3
 Query Match: 58.54% Indels: 0
 DB: 2 Gaps: 0
 US-10-081-935-2 (1-18) x AC130471 (1-148008)

```

Oy 1 MetGlnPheSerHisAlaLeuValAlaLeuValAlaGlyLeuAlaSer 17
Db 47008 CTCCAATTACAGGATATTACTGACTTCTAGTATCTGCAGGAATGGCCAGT 47058

RESULT 16
AC011421 148142 bp DNA linear PRI 18-APR-2000
LOCUS Homo sapiens chromosome 5 clone CTD-230305, complete sequence.
AC011421
AC011421.2 GI:6102634
HTG:
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 148142)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 148142)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 148142)
AUTHORS DOE Joint Genome Center and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1999) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 148142)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 5 (bases 1 to 148142)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 6 (bases 1 to 148142)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Oct 23, 1999 this sequence version replaced gi:6013521.
Draft Sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
STS Content:
WI-6744 G05721
WI-6988 G06342.
Location/Qualifiers
1. 148142
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-230305"
BASE COUNT 50576 a 26865 c 24848 g 45853 t
ORIGIN

Alignment Scores:
Pred. No.: 3.31e+03 Length: 148142
Score: 48.00 Matches: 9
Percent Similarity: 82.35% Conservative: 5
Best Local Similarity: 52.94% Mismatches: 3
Query Match: 58.54% Indels: 0
DB: 9 Gaps: 0

US-10-081-935-2 (1-18) x AC011421 (1-148142)
Oy 1 MetGlnPheSerHisAlaLeuValAlaLeuValAlaGlyLeuAlaSer 17
Db 92942 CTCCAATTACAGGATATTACTGACTTCTAGTATCTGCAGGAATGGCCAGT 92992

RESULT 17
AC011783 175178 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-16L10, LOW-PASS SEQUENCE SAMPLING.
AC011783
AC011783.3 GI:7144945
HTG: HTGS PHASEO.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 175178)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Unpublished
JOURNAL Homo sapiens, clone RP11-16L10
REFERENCE 2 (bases 1 to 175178)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Batna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Gallagan, J., Gardy, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Kartas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 3, 2000 this sequence version replaced gi:6446867.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3568
Center Clone name: 16_L_10
-----
* NOTE: This record contains 185 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 916: contig of 916 bp in length
* 917 1016: gap of 100 bp
* 1017 1893: contig of 877 bp in length
* 1894 1993: gap of 100 bp
* 1994 2899: contig of 906 bp in length
* 2900 2999: gap of 100 bp
* 3000 3975: contig of 976 bp in length
* 3976 4075: gap of 100 bp
* 4076 5009: contig of 934 bp in length
* 5010 5109: gap of 100 bp

```

5110	6049: contig of 940 bp in length	42919	43018: gap of 100 bp
6050	6149: gap of 100 bp	43019	43943: contig of 925 bp in length
6150	7075: contig of 926 bp in length	43944	44043: gap of 100 bp
7076	7175: gap of 100 bp	44044	44994: contig of 951 bp in length
8110	8110: contig of 935 bp in length	44995	45094: gap of 100 bp
8211	8210: gap of 100 bp	45095	46032: contig of 938 bp in length
8211	9126: contig of 916 bp in length	46033	46132: gap of 100 bp
9127	9226: gap of 100 bp	46133	47063: contig of 931 bp in length
9227	10157: contig of 931 bp in length	47064	47163: gap of 100 bp
10158	10257: gap of 100 bp	47164	48113: contig of 950 bp in length
10258	11178: contig of 921 bp in length	48114	48213: gap of 100 bp
11179	11278: gap of 100 bp	48214	49162: contig of 949 bp in length
11279	12202: contig of 924 bp in length	49163	49262: gap of 100 bp
12203	12302: gap of 100 bp	49263	50166: contig of 904 bp in length
12303	13218: contig of 916 bp in length	50167	50266: gap of 100 bp
13219	13318: gap of 100 bp	50267	51189: contig of 923 bp in length
13319	14252: contig of 934 bp in length	51190	51289: gap of 100 bp
14253	14352: gap of 100 bp	51290	52212: contig of 923 bp in length
14353	15290: contig of 938 bp in length	52213	52312: gap of 100 bp
15291	15390: gap of 100 bp	52313	53209: contig of 897 bp in length
15391	16319: contig of 929 bp in length	53210	53309: gap of 100 bp
16320	16419: gap of 100 bp	53310	54254: contig of 945 bp in length
16420	17356: contig of 937 bp in length	54255	54354: gap of 100 bp
17357	17456: gap of 100 bp	54355	55298: contig of 944 bp in length
17457	18374: contig of 918 bp in length	55299	55398: gap of 100 bp
18375	18374: gap of 100 bp	55399	56307: contig of 909 bp in length
18475	19399: contig of 925 bp in length	56308	56407: gap of 100 bp
19400	19499: gap of 100 bp	56408	57325: contig of 918 bp in length
19500	20434: contig of 935 bp in length	57326	57425: gap of 100 bp
20435	20534: gap of 100 bp	57426	58366: contig of 941 bp in length
20535	21209: contig of 675 bp in length	58367	58466: gap of 100 bp
21210	21309: gap of 100 bp	58467	59390: contig of 924 bp in length
21310	22235: contig of 926 bp in length	59391	59490: gap of 100 bp
22236	22335: gap of 100 bp	59491	60416: contig of 926 bp in length
22336	22993: contig of 958 bp in length	60417	60516: gap of 100 bp
22994	23393: gap of 100 bp	60517	61462: contig of 946 bp in length
23394	24333: contig of 940 bp in length	61463	61562: gap of 100 bp
24334	24433: gap of 100 bp	61563	62510: contig of 948 bp in length
24434	23365: contig of 932 bp in length	62511	62610: gap of 100 bp
25366	25465: gap of 100 bp	62611	63538: contig of 928 bp in length
25466	26394: contig of 929 bp in length	63539	63638: gap of 100 bp
26395	26494: gap of 100 bp	63639	64574: contig of 936 bp in length
26495	27441: contig of 947 bp in length	64575	64674: gap of 100 bp
27442	27541: gap of 100 bp	64675	65613: contig of 939 bp in length
27542	28440: contig of 899 bp in length	65614	65713: gap of 100 bp
28441	28540: gap of 100 bp	65714	66625: contig of 912 bp in length
28541	29468: contig of 928 bp in length	66626	66725: gap of 100 bp
29469	29568: gap of 100 bp	66726	67336: contig of 911 bp in length
29569	30504: contig of 936 bp in length	67337	67736: gap of 100 bp
30505	30604: gap of 100 bp	67737	68689: contig of 953 bp in length
30605	31513: contig of 909 bp in length	68690	68789: gap of 100 bp
31514	31613: gap of 100 bp	68790	69734: contig of 945 bp in length
31614	32552: contig of 939 bp in length	69735	69834: gap of 100 bp
32553	32652: gap of 100 bp	69835	70735: contig of 901 bp in length
32653	33586: contig of 934 bp in length	70736	70835: gap of 100 bp
33587	33686: gap of 100 bp	70836	71771: contig of 936 bp in length
33687	34615: contig of 929 bp in length	71772	71871: gap of 100 bp
34616	34715: gap of 100 bp	71872	72819: contig of 948 bp in length
34716	35640: contig of 925 bp in length	72820	72919: gap of 100 bp
35641	35740: gap of 100 bp	72920	73902: contig of 983 bp in length
35741	36676: contig of 936 bp in length	73903	74002: gap of 100 bp
36677	36776: gap of 100 bp	74003	74924: contig of 922 bp in length
36777	37684: contig of 908 bp in length		
37685	37784: gap of 100 bp		
37785	38781: contig of 937 bp in length		
38782	38881: gap of 100 bp		
38882	39829: contig of 948 bp in length		
39829	39929: gap of 100 bp		
39930	40857: contig of 928 bp in length		
40858	40957: gap of 100 bp		
40958	41882: contig of 925 bp in length		
41883	41982: gap of 100 bp		
41983	42918: contig of 936 bp in length		

Alignment Scores:
 Pred. No.: 3.93e+03 Length: 175178
 Score: 48.00 Matches: 9
 Percent Similarity: 82.35% Conservative: 5
 Best Local Similarity: 52.94% Mismatches: 3
 Query Match: 58.54% Indels: 0
 DB: 2 Gaps: 0

US-10-081-935-2 (1-18) x AC011783 (1-175178)

```

QY      1 MetGlnPheSerHisAlaLeuValAlaLeuValAlaAlaGlyLeuAlaSer 17
Db      44746 CTCCAATTACAGGATATTACTGACTTTAGTATCGCAGGAATGCCAGT 44696

RESULT 18
LOCUS   GGA252171/c
DEFINITION Gallus gallus partial mRNA for proprotein convertase PACE4 (PACE4 gene), clone cHPACE.
ACCESSION AJ252171
VERSION AJ252171.1 GI:6706187
KEYWORDS PACE4 gene; proprotein convertase PACE4.
SOURCE Gallus gallus (Chicken)
ORGANISM Gallus gallus

REFERENCE
AUTHORS Feldmann,A., Schafer,M.K., Garten,W. and Klenk,H.D.
TITLE Targeted infection of endothelial cells by avian influenza virus A/FPV/Rostock/34 (H7N1) in chicken embryos
JOURNAL J. Virol. 74 (17), 8018-8027 (2000)
MEDLINE 20392157
PUBMED 10933711
REFERENCE 2 (bases 1 to 377)
AUTHORS Feldmann,A.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2000) Feldmann A., Institut fuer Virologie, Philipps Universitaet Marburg, Robert-Koch-Str. 17, 35037 Marburg, GERMANY

FEATURES
Source
Location/Qualifiers
1..377
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="cHPACE"
/tissue_type="liver"
/dev_stage="11-day-old"
/country="Germany"
1..377
/gene="PACE4"
<1..>377
/gene="PACE4"
/codon_start=1
/product="proprotein convertase PACE4"
/protein_id="CAB65907.1"
/db_xref="GI:6706188"
/db_xref="SPTREMBL:Q9PT85"
/translation="HKVSHFYGFLGLDAEAIIVEAKKWKTPPOHVCVCSLDRVPKYI
RPDHLVRLASTLSSACSEOEQHVLYLHHVVVRLSIHPRGRDLQISLVSPAGTRSQLL
ARRVFDHNEGFGKGEFMTVHCWG"

BASE COUNT 80 a 107 c 122 g 68 t
ORIGIN

Alignment Scores:
Pred. No.: 11.9 Length: 377
Score: 47.00 Matches: 9
Percent Similarity: 72.22% Conservative: 4
Best Local Similarity: 50.00% Mismatches: 5
Query Match: 57.32% Indels: 0
DB: 5 Gaps: 0

US-10-081-935-2 (1-18) x GGA252171 (1-377)

QY      1 MetGlnPheSerHisAlaLeuValAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db      213 CTCACAGTACAGCAGCTGCTCGCTGCTCGCTCGCTCGCAGGACCTGCTCAGAGT 160

RESULT 19
LOCUS   AE006992
DEFINITION Mycobacterium tuberculosis CDC1551, section 78 of 280 of the complete genome.

AE006992
LOCUS   AE006992
DEFINITION Mycobacterium tuberculosis CDC1551, section 78 of 280 of the complete genome.

```

```

ACCESSION AE006992 AE000516
VERSION AE006992.1 GI:13880691
KEYWORDS
SOURCE
ORGANISM Mycobacterium tuberculosis CDC1551
Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteria; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 14483)
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14483)
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H., Gill,J., Mikula,A. and Bishai,W.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Source
Location/Qualifiers
1..14483
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
/complement(66..560)
/gene="MT1111"
/complement(66..560)
/gene="MT1111"
/note="similar to SP:P21346; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcription elongation factor GreA"
/protein_id="AAK45367.1"
/db_xref="GI:13880692"
/translation="MTDQVTWLTOESHDLKAEILDQIANRPVIAEINDRREGDG
RENGYHAAAREEQGOEARIQOLLSNAKVGEA*QSGVALPGSVKVVYNGDSD
SETFLIATROEGVSDGKLEVYSPNSPLGALLIDAKVGETRSTVTPNGSTVSVTLVSAE
PVHS"
/complement(746..1207)
/gene="MT1112"
/complement(746..1207)
/gene="MT1112"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45368.1"
/db_xref="GI:13880693"
/translation="MHPAGVNPCTHTPIPRDARYGRPLRRRRRVAIALGLVLA
AAGIVIAVIGTQRISTASVGLVRLVDDETASVTISVTRSDSPRPVACIVRVRA
NGSETGRRELIVPPSEATTQVTTTKVSKQPPVADVYCGGTEVPSYLRP"
1282..2148
/gene="MT1113"
1282..2148
/gene="MT1113"
/note="similar to PID:699274; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="lmbE protein"
/protein_id="AAK45369.1"
/db_xref="GI:13880694"
/translation="MSELRLMVAHVAHPDDESSKGAATLARYADEGHRVLVVTLTGGER

```



```

SOURCE
ORGANISM      Mycobacterium tuberculosis H37Rv
Bacteria; Actinobacteria; Actinomycetales;
Corynebacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
AUTHORS      1
              Cole, S. T., Brosch, P., Parkhill, J., Garnier, T., Churcher, C.,
              Harris, D., Gordon, S. V., Eigmeier, K., Gas, S., Barry III, C. E.,
              Tekai, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
              Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
              Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
              Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M. A.,
              Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
              Squares, S., Squires, R., Sulston, J. E., Taylor, K., Whitehead, S. and
              Barrell, B. G.
TITLE         Deciphering the biology of Mycobacterium tuberculosis from the
JOURNAL       Nature 393 (6685), 537-544 (1998)
MEDLINE       98295987
PUBMED        9634230
REFERENCE     2 (bases 1 to 67200)
AUTHORS      Parkhill, J.
TITLE         Direct Submission
JOURNAL       Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
              tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
              Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
              Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
              75124 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
              On Jun 26, 1998 this sequence version replaced gi:2896684.
COMMENT       Notes:
              Details of M. tuberculosis sequencing at the Sanger Centre are
              available on the World Wide Web.
              (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
              been renumbered from the original cosmid submissions but the old
              gene designations are in brackets after the new gene numbers.
              Gene predictions are in brackets after the new gene numbers.
              implemented in TParse (Krogh) supplemented with visual inspection
              of positional base preference in codons, especially where there is
              an increase in the observed/expected third position G + C.
              CAUTION: In some cases we may not have predicted the correct
              initiation codon. Where possible we choose an initiation codon
              (atg, gtc, or ttc) which is preceded by an upstream ribosome
              binding site sequence (optimally 5-13bp before the initiation
              codon). If this cannot be identified we choose the most upstream
              initiation codon.
              Location/Qualifiers
              1..67200
              /organism="Mycobacterium tuberculosis H37Rv"
              /mol_type="genomic DNA"
              /strain="H37Rv"
              /db_xref="taxon:83332"
              66070..>67200
              /organism="Mycobacterium tuberculosis H37Rv"
              /mol_type="genomic DNA"
              /strain="H37Rv"
              /db_xref="taxon:83332"
              /clone="12268"
              <1..66170
              /note="fragment designated v017. Does not represent a
              physical clone"
              178..181
              /note="possible RBS, ggaag, for Rv1049"
              184..630
              /gene="Rv1049"
              184..630
              /gene="Rv1049"
              /note="Rv1049", (MTV017.02), len: 148. Probable repressor
              similar to many e.g. P74870 (149 aa). FASTA scores:
              sp|P74870|P74870NEGATIVE REGULATORY OF EMR LOCUS EMR (149
              aa) opt:146 z-score: 200.3 E(1): 0.0011; 31.6% identity in
              95 aa overlap. TParse score is 0.892. Contains probable
              helix-turn-helix motif at aa 58-79 (Score 1495, +4.28 SD)
              /codon_start=1

FEATURES
source
1..67200
/organism="Mycobacterium tuberculosis H37Rv"
/mol_type="genomic DNA"
/strain="H37Rv"
/db_xref="taxon:83332"
66070..>67200
/organism="Mycobacterium tuberculosis H37Rv"
/mol_type="genomic DNA"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="12268"
<1..66170
/note="fragment designated v017. Does not represent a
physical clone"
178..181
/note="possible RBS, ggaag, for Rv1049"
184..630
/gene="Rv1049"
184..630
/gene="Rv1049"
/note="Rv1049", (MTV017.02), len: 148. Probable repressor
similar to many e.g. P74870 (149 aa). FASTA scores:
sp|P74870|P74870NEGATIVE REGULATORY OF EMR LOCUS EMR (149
aa) opt:146 z-score: 200.3 E(1): 0.0011; 31.6% identity in
95 aa overlap. TParse score is 0.892. Contains probable
helix-turn-helix motif at aa 58-79 (Score 1495, +4.28 SD)
/codon_start=1

gene
CDS
/translation="MGKGAFFDEACACCTTTRRAAROLGQAYDRALRPSGLTNTQSTLA
VLSSEGAGIDLTMSLELARIQVETTLTNLEVRMDGVRVWAGADARCKRIELT
AGKGAALOKAVPLMRGQAEVTASVGDMPRRRDINLGOAAEACR"
668..672
/note="possible RBS, aagag, for Rv1050"
679..1584
/gene="Rv1050"
679..1584
/gene="Rv1050"
/note="Rv1050", (MTV017.03), len: 301. Probable
oxidoreductase similar to many e.g. MCV48.22C(341 aa).
FASTA scores: sp|O10783|Y04M_MYCTU_PUTATIVE OXIDOREDUCTASE
CY48.22C (341aa) opt:462 z-score: 533.9 E(1): 3e-22; 33.6%
identity in265 aa overlap. TParse score is 0.896"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1050"
/protein_id="CAAI1716.1"
/db_xref="GI:2896687"
/db_xref="SPTREMBL:O53398"
/translation="MAROPRFDVVLITGASGGIGFATKAFAREGAVLAAAREGA
LRVAREIAAGGAWAPLDVSSSEVRANVAVNGEFGIDVFNAGVSLVCPVD
AETFLDTRMLELIDYGTAVVAREVPLMKQSGSRIMNNSVVGRAFARFAYSS
AMHAIAGFSDALRQELRSGIAGVAVIHPLATQPLANVADPADMPPPRSLLPIPVH
VAAVLDGVARRARVAVVPPQPLMLMGDAFSPRYCDRVVRLLESKIFGRUGSYRGS
VYRHPTESAKQAQAQPERGYSAR"
complement (11743..2498)
/gene="Rv1051c"
complement (11743..2498)
/gene="Rv1051c"
/translation="MRADVTAEHLTOVVRDIAVIDIDDVAFNLDTSVQVIRERADY
RGLRVVAMSGVWQCIAMDVSTGEPIAAMPTRVTRDILGEFTITLGTAFETIAB
KGVTLIERGTTSTRMDYVDIVDRLRGIDDDLELRSARVAVOQRGATLEVAATHLAG
YGAIVAQKAWATERGQCHWRHKPAVWGRNMDLDAKQVSEMIQVPGVTLRHMRRS
DIGPASTFLGRVVRVYRDEVSRWISKESATRR"
3507..3513
/note="possible RBS, aaggtgg, for Rv1052"
3521..3910
/gene="Rv1052"
3521..3910
/gene="Rv1052"
/note="Rv1052", (MTV017.05), len: 129. Unknown"
/codon_start=1
/transl_table=1
/product="hypothetical protein Rv1052"
/protein_id="CAAI1716.1"
/db_xref="GI:2896689"
/db_xref="SPTREMBL:O53400"
/translation="MDCCEERGVARHKGISQVGTGCCPRMSOAVSVCASVYRAAYTA
VQWPLRPGVEPLPHDETLAALLPVVEVDKRTITADVDYDLEQGLQDOVFDLMLPTA
VBSLSLDBLSHFRFDRLHARHFGV"
complement (13809..4084)
/gene="Rv1053c"
complement (13809..4084)

```

```

gene
/note="RV1053c" (MTV017.06c), len: 91. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV1053c"
/protein_id="CA11710.1"
/db_xref="GI:2896690"
/translation="MDSHKVCNMNNNLTPTGPIIGHPAPVADGVRVAYLUDGLLRNC"
TDVFTSSPPGPVLVYRKTRVLEADMTVEKLKIKORANSRRHQ"
4726..4932
/gene="RV1054"
4726..>4932
/gene="RV1054"
/note="RV1054. (MTV017.07), len: 68. Unknown but similar
to hypothetical Mycobacterium tuberculosis protein
MTCY3G12_25 (151 aa shows similarity to integrases) and tcc
Mycobacterium paratuberculosis integrase MScint_1 (191
aa). This orf continues in another frame as MTV017.08 but not
enough to be found to account for frameshift. FASTA scores:
279702|MTCY3G12_25 (151 aa) opt: 273 z-score: 364.1 E():
8.8e-13; 64.7% identity in 68 aa overlap; and
L39071|MSGINT_1 (191aa)opt: 105 z-score: 148.5 E(): 0.9;
31.8% identity in85 aaoverlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein RV1054"
/protein_id="CA11710.1"
/db_xref="GI:2896691"
/db_xref="SPTREMBL:O533402"
/translation="WTGKGIVESTTKTRDNRHVPEVWRRLHAELETPDNLVFPQC
RKGGFLPGEYRAMPFDNAGDVGIE"
4935..5171
/gene="RV1055"
4935..5171
/gene="RV1055"
/note="RV1055. (MTV017.08), len: 78. Partial orf, first
49aa similar to hypothetical Mycobacterium tuberculosis
protein MTCY3G12_25 (151 aa shows similarity to
integrases) and to Mycobacterium paratuberculosis integrase
L39071|MSGINT_1 (191 aa) and to many other integrases or
transposases. FASTA scores: 279702|MTCY3G12_25 (151 aa) opt:
291 z-score: 428.6 E(): 2.2e-16; 74.3% identity in 70 aa
overlap; and gpL39071|MSGINT_1 (191 aa) opt: 146 z-score:
221.0 E(): 8.3e-05; 52.1% identity in 48 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV1055"
/protein_id="CA11711.1"

Alignment Scores:
Pred. No.: 2.3e+03 Length: 67200
Score: 47.00 Matches: 9
Percent Similarity: 92.86% Conservative: 4
Best Local Similarity: 54.29% Mismatches: 1
Query Match: 57.32% Indels: 0
DB: 1 Gaps: 0

US-10-081-935-2 (1-18) x MTV017 (1-67200)

QY 3 PheserHISAlaLeuIleAlaLeuValAlaIaIaGlyLeuAla 16
|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 46872 TTCCGCGACGCCCTGCTGCATGCTTCCGCGTGAATTGGCG 46913

RESULT 21
AC144792 210652 bp DNA linear HTG 20-MAY-2003
LOCUS AC144792 Mus musculus chromosome UNK clone RP24-501M24, WORKING DRAFT
DEFINITION AC144792.1 GI:30911145
SEQUENCE, 4 unordered pieces.
AC144792.1 GI:30911145
AC144792.1 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus (house mouse)
SOURCE

```

```

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
Wilson,R.K.
1 (bases 1 to 210652)

JOURNAL
TITLE
The sequence of Mus musculus clone
Unpublished
REFERENCE
2 (bases 1 to 210652)
AUTHORS
Wilson,R.K.
Direct Submission
Submitted (20-MAY-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
-----
Center project name: M BB0501M24

----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: Plasmid, 100%
Chemistry: Dye-Primer ET, 0% of reads
Chemistry: Dye-Terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 209398 bases at least Q40
Consensus quality: 209715 bases at least Q10
Consensus quality: 209918 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 12335: contig of 12335 bp in length
* 12336 12435: gap of unknown length
* 12436 37833: contig of 25398 bp in length
* 37834 37933: gap of unknown length
* 37934 94620: contig of 56687 bp in length
* 94621 94720: gap of unknown length
* 94721 210652: contig of 115932 bp in length.
*
Location/Qualifiers
1. .210652
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-501M24"
1. .12335
/note="assembly_name:Contig17"
12436. 37833
/note="assembly_name:Contig18"
37934. 94620
/note="assembly_name:Contig19"
94721. 210652
/note="assembly_name:Contig20"

BASE COUNT 54629 a 48568 c 48742 g 58398 t 315 others
ORIGIN

Alignment Scores:
Pred. No.: 7.35e+03 Length: 210652
Score: 47.00 Matches: 8
Percent Similarity: 82.35% Conservative: 6
Best Local Similarity: 47.06% Mismatches: 3
Query Match: 57.32% Indels: 0
DB: 2 Gaps: 0

US-10-081-935-2 (1-18) x AC144792 (1-210652)

```

Cy I MetcnpHeserrHisAlaLeuIleAlaLeuValAlaAacGlyLeuAlaSer 17
Db 159347 TTGAAGTTTGACATCACCCTTCGTGCATGGTGCATCGAGCATGTGAGGC 159397

RESULT 22

BX248337
LOCUS BX248337
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 4/14.
ACCESSION BX248337 BX248333
VERSION BX248337.1 GI:31617663
KEYWORDS complete genome.
SOURCE Mycobacterium bovis subsp. bovis AF2122/97
ORGANISM Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinomycetidae; Actinomycetales;
Corynebacterales; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE
AUTHORS
1 Garnier,T., Eiglmeyer,K., Camus,J.-C., Medina,N., Mansoor,H., Piryot,M., Durloy,S., Grondin,S., Lacroix,C., Monsempé,C., Simon,S., Harris,B., Aklin,R., Doggett,J., Mayes,R., Keating,L., Wheeler,P.R., Parkhill,J., Barrrell,B.G., Cole,S.T., Gordon,S.V. and Hewinson,G.
The complete genome sequence of Mycobacterium bovis Online Publication PMID: 10.1073/pnas.1130426100 (Microbiology)
2 (bases 1 to 327650)
Garnier,T.
Direct Submission
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Molculaire Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex 15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of The Mycobacteries Agency sequencing teams, TB Research Group, Veterinary Laboratories Agency Weybridge, Woodham Lane, New Haw,Addlestone, Surrey KT15 3NB, UK, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,Cambridge CB10 1SA, UK, PT4 Annotation, Genopole, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Molculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France
Location/Qualifiers

FEATURES
SOURCE
1..327650
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
698..2170
/locus_tag="Mb0912"
698..2170
/locus_tag="Mb0912"
/notes="Mb0912." len: 490 aa. Equivalent to Rv0888, len: 490 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 490 aa overlap). Probable exported protein. Equivalent to AAK5157.1 from Mycobacterium tuberculosis strain CDC1551 (507 aa) but shorter 17 aa. contains possible N-terminal signal sequence."
/codon_start=1
/transl_table=1
/product="PROBABILE EXPORTED PROTEIN"
/protein_id="CAD93773.1"
/db_xref="GI:31617664"
translation="MDYAKRGVGVGLAVLVGVAATTTHASAPTDSSSTDP VDACPISGSASSLSAIKPGASVPQVGROUDPSIPDDLNALIDFLAAVRNGIVPI ENRPYPVANPOOVSPVEEGTVGPVFADCDPDDGNMTFAVERERGPGSGHGITVDOR TASVTIADPELFTDIFSVNVSDDTSLHVHAGLAGLYGPHGHDAVVTFVGNPT DTJDSGFMLTYNVIAGLPFPLSSAILPRFFYTEKGIGRLAYVAVNAOEPAHQFLLIKKSMPSETPPEPTLMILGVPFGDLINTISEKRVLRQWYECSNTCLKGE TYSMRNLPGGITDVVNLIHTNYGGSPPTNNLAADVANIIONSAGRATIVTGPNARY SDOSALAEFOAVNGLDIAWOVQEHGRTTPEPATCNMGECLELDKITFRSGCVTL OAVSYENAEERKFENSKEEPLSDHSFPAVVGHHYADNAVNR"
complement(1205)..3326
/gene="CitCa"
/locus_tag="Mb0913c"

complement(2205, .3326)
/gene="citA"
/locus_tag="Mb0913c"
/EC_number="2.3.3.1"
/note="Mb0913c, citA, len: 373 aa, from Mycobacterium tuberculosis strain H37Rv (100.0% identity in 373 aa overlap). Probable citA (alternate gene name: gltA), citrate synthase 2 (EC 4.1.3.7), highly similar to others e.g. CAB95899.1|AL359988 putative citrate synthase from Streptomyces coelicolor (387 aa); P39119|C15Y_BACSU citrate synthase II from Bacillus subtilis (366 aa), FASTA scores: opt: 586, E(): 5.8e-30, (33.8% identity in 367 aa overlap), etc. Also similar to Rv0896|MT031.24 from Mycobacterium tuberculosis (29.2% identity in 274 aa overlap) and Rv1131. Contains P500480 Citrate synthase signature. BELONGS TO THE CITRATE SYNTHASE FAMILY."
/codon_start=1
/transl_table=1
/product="PROBABLE CITRATE SYNTHASE II CITA"
/protein_id="CAD93774.1"
/db_xref="GI:31617665"
/translation="MTVPENFVPLGDGVAFETTEIAEPDGDGALRYGVIEDLVLE
ORVTEGDMALVDGNFGSGCLPAPPEFLPHSGCVRVDVQAGLMLPIGAYALL
IDTAPRQOLKASVMAVSVAOASRGLYGVAPVPR1IDESTVAPRMTMGCPD
RHIEDIDVYVWVSAEHCWMASTFTARVAVASGADVAALASICAGMSGLPGARPA
LPMDEVERRAGDASVVKGILDRGEKLMGFRRYRYRADPPARVULRAAEIKGAPRY
VAVAEQALSKLVEERRPRDRA1ETNVEFMAAVVDVFAVPANMMPAMFTCGRTAGWC
H1EEKRGKGLKVPRAIPVAPGPRSPESVDGMERVLTTA"
complement(3413, .6061)
/locus_tag="Mb0914c"
/locus_tag="Mb0914c"
complement(3413, .6061)
/locus_tag="Mb0914c"
/note="Mb0914c, -, len: 882 aa. Equivalent to Rv0890C, len: 882 aa, from Mycobacterium tuberculosis strain H37Rv, (99.9% identity in 882 aa overlap). Probable transcriptional regulatory protein, LuxR family, highly similar (but shorter 238 aa in N-terminus) to NP_302202.1|NC_002677 possible transcriptional regulator from Mycobacterium lepre (1106 aa). Also highly similar (generally in part) to others e.g. T50568 probable multi-domain regulatory protein from Streptomyces coelicolor (1334 aa); P10957|NMRL_ECOLI nitrate/nitrite response regulator protein from Escherichia coli (216 aa) FASTA scores: opt: 193, E(): 6e-06, (37.4% identity in 99 aa overlap), etc. Also highly similar to others from Mycobacterium tuberculosis e.g. MYC02B10_22, MYV008_44, MYV036_21, and MYC31_24. Contains P50017 ATP/GTP-binding site motif A (P-loop), P500622 Bacterial regulatory proteins, LuxR family signature, and probable helix-turn helix motif from aa 836 to 857 (Score 1559, +4.50 SD). BELONGS TO THE LUXR/HMPA FAMILY OF TRANSCRIPTIONAL REGULATORS."
/codon_start=1
/transl_table=1
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN (PROBABLY LUXR-FAMILY)"
/protein_id="CAD93775.1"
/db_xref="GI:31617666"
/translation="MRALLAONRLVLTGTCGVGKTRLA1DIASASELRDGLCFVDL
ETISGIVAAAPRAVAGLPDQGRSTWDSLRPIFGNRMWLVNCEHLLDACALVAV
ELTACPELTLATSRPTGAGC1YTRVPSMTIDEVLE1FAPRASVVGFTIANH
NAAVAGEICRLDGIPLA1EFAARVMSPLE1ADKMTLCOTLYRGVGTCSGDPI
LRASIDMSHLLDETGOILFRLRAPVCGFLAVRAVAAASDIDPSVLDLIVD
KSLVADDCOGRVRLLETYRVALCEGSDGADVAHRRVYDLSASLNPAN
DHQULVARETEINDLRAAPFMSRENKHITBALDASLOPI1RGRHLREGLSFNPS
ILBOGRHRLR1VSTRARALADPAM1STWAT1SVGAT1D1IAAQOLAARVAGD
AALVRAAECCDCCSSGVNAEAA1PYFAASCTIDLA1DKMTLCOTLYRGVGTCSGD
NALVRAAECCDCCSSGVNAEAA1PYFAASCTIDLA1DKMTLCOTLYRGVGTCSGD
DVPTVGLVYQAOVLVACGASAAH1AGACT1AATETAGVYCG1GYAATVYALAG
DPTVALSEDAAR1ILRAOPDVYTMHVLVQALAGDGI1AAROPANDVATNGMH
RMVAT1LRAVATRGEBELARDHAA1ACGA1LH1YCCPMDE1LAGAGVSGH
SEGR1LGAAT1AAROPDVRVYVNCVQ1STH1LR1WPRDREMLN1AGTCSGSH

gene
complement (6063...6920)
/locus_tag="Mb0915c"
/locus_tag="Mb0915c"
/note="Mb0915c", len: 285 aa. Equivalent to Rv0891C,
len: 285 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.6% identity in 285 aa overlap). Possible
transcriptional regulator, highly similar in N-terminus to
NP_302202.1|NC_002677 possible transcriptional regulator
from Mycobacterium leprae (1106 aa). Also highly similar
to several Mycobacterium tuberculosis putative
transcriptional regulators e.g. Q1102|MT02B10_22
PROBABLE TRANSCRIPTIONAL REGULATOR PROTEIN (1159 aa),
FASTA scores: opt: 702, E(1): 8.3e-40, (50.6% identity in
247 aa overlap); MT0036_21; MT0008_44; MT002B10_23. Also
shows similarity with several adenylate cyclases and
hydrolases from other organisms."

gene
CDS
/transl_table=1
/product="POSSIBLE TRANSCRIPTIONAL REGULATORY PROTEIN"
/protein_id="CAD93776.1"
/db_xref="GI:31617667"
/translation="MLFNAVNSLPNIDIDHAIIRGEDHPPTCAKCVAGRISALGS
LDLRHSRCYAAPPPDYRCFPPRRRVLIANOGLDVSRLPPTGTTLLADVEEST
HLMQCPEDMATIAHLDHTVEAITHHGVQPVKRYEGSFVAFTASDAACALD
LQRTSLAPRLRLGLHTGEVOLRDELVLVPTINTARILRHAGQVLSLATDLYT
GRLPADMLVDLGRHPLRGLPRPVMVQLCPHIREKPEPLRTAKSSPTSLPAQFTT
FVGRRAQIS"
7318...8805
/locus_tag="Mb0916"
7318...8805
/locus_tag="Mb0916"
/EC_number="1.14.-.-"
/note="Mb0916", len: 495 aa. Equivalent to Rv0892, len:
495 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 495 aa overlap). Probable
monooxygenase (EC 1.14.-.-), highly similar to others e.g.
NP_250787.1|NC_002516 probable flavin-binding
monooxygenase from Pseudomonas aeruginosa (491 aa);
CAB59668.1|AL132674 monooxygenase from Streptomyces
coelicolor (519 aa); P12015|CYMO ACIS cyclohexanone
monooxygenase from Acinetobacter sp. (542 aa), FASTA
scores: opt: 489, E(1): 6.8e-26, (30.3% identity in 492 aa
overlap); etc. Also highly similar to Rv0565c, Rv3854c,
hydrophobic stretch at N-terminus."
/codon_start=1
/transl_table=1
/product="PROBABLE MONOOXYGENASE"
/protein_id="CAD93777.1"
/db_xref="GI:31617668"
/translation="MTGRCPYAVAVGAGSMCAVITLLSAGITDVCIEKADVGST
WRDNTYGLTCDVPSRLYOXFAKNPMWTFSGGSGIOYLRGIRARIRRG
ATVVSARFDDGRWLRTDSCGSEVDFLLIATGVLHPRIPRAGADDPGCTHSAR
WDTHTPLIGRIAVIGCTGTCVUCLAGACVATKFTQAVLDPWPPRSKLAR
VFRAPFCGLSLAVKAYSLEPFAVALSNPGLHRKLVGAVCRASLRVSDPRRLAL
TPDYEPWCKRLVMSGSGFYRAIQRDVGLVAGIDHVEHGIIVTDGVLHVDVIVLAT
GFDSHAFFRPMOLTGDRIRIDVMDQGPAAHQVAIPGFNPFMGLPSPVGNFPL
TAAVESQAEHIVQMIKRRHGEFDTMEPKSAATEAVYTVLRAMPNTVMTGDSWYL

Alignment Scores:
Pred. No.: 1.15e+04 Length: 327650
Score: 47.00 Matches: 9
Percent Similarity: 92.86% Conservative: 4
Best Local Similarity: 64.29% Mismatches: 1
Query Match: 57.32% Indels: 0
DB: 1 Gaps: 0

US-10-081-935-2 (1-18) x BX248337 (1-327650)
QY 3 PheserHisAlaLeuValAlaAlaClyLeuAla 16

Db 233446 TTGGGACGCGCCCTGTGTCATGCTTGGCCCTGAATTGGC 233487
RESULT 23
G32057
LOCUS MCM301 Chicken RPMACrooijmans Gallus gallus STS genomic, sequence
DEFINITION tagged site.
ACCESSION G32057.1 GI:2138003
VERSION G32057.1
KEYWORDS STS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 1204)
Crooijmans R.P., Dijkhof R.J., van der Poel J.J. and Groenen M.A.
New microsatellite markers in chicken optimized for automated
fluorescent genotyping
Anim. Genet. 28 (6), 427-437 (1997)
JOURNAL
MEDLINE 98251489
PUBMED 9589584
COMMENT
Contact: Richard Crooijmans
Department of Animal Breeding
Wageningen Agricultural University
Box 338, 6700 AH Wageningen, The Netherlands
Email: richard.crooijmans@alg.vf.war.nl
Primer A: GGAGTACGACCACTGATTC
Primer B: AGGCTAGAGTACCAAGTGC
STS size: 278
PCR Profile:
Presoak: 94 degrees C for 5 minute (s) :
Denaturation: 94 degrees C for 30 seconds
Annealing: 55 degrees C for 45 seconds
Polymerization: 72 degrees C for 90 seconds
PCR Cycles: 35
Thermal Cycler: MJ research
Protocol:
Template: 10-100 ng
Primer: each 295 nM
dNTPs: each 200 uM
Polymerase Goldstar: 0.4 u
Total Vol: 12 kl
Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
Tetramethylammoniumchloride (TMAc): 1 mM
pH: 8.3

FEATURES
source size range: 266-284
Location/Qualifiers
1..1204
/organism="Gallus gallus"
/mol_type="genomic DNA"
/db_xref="taxon:9031"
/clone_1lb="Chicken RPMACrooijmans"
STS
Primer_bind 290..567
primer_bind 290..310
BASE COUNT 260 a 279 c 294 g 360 t 11 others
ORIGIN

Alignment Scores:
Pred. No.: 60 Length: 1204
Score: 46.00 Matches: 9
Percent Similarity: 92.31% Conservative: 3
Best Local Similarity: 69.23% Mismatches: 1
Query Match: 56.10% Indels: 0
DB: 11 Gaps: 0

Score: 46.00 Matches: 10
Percent Similarity: 84.62% Conservative: 1
Best Local Similarity: 76.92% Mismatches: 2
Query Match: 56.10% Indels: 0
Gaps: 0
DB: 8

US-10-081-935-2 (1-18) x AY036106 (1-4029)
QY 5 HisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSer 17
Db 2487 CACGACCTTGACGCCCTTGCTGCTGACTGACGCC 2449

RESULT 26
LOCUS MMU307017 8117 bp mRNA linear ROD 14-MAY-2001
DEFINITION Mus musculus mRNA for putative ubiquitin-specific protease (Uspsy
gene).

ACCESSION AJ307017 GI:14041781
VERSION AJ307017.1
KEYWORDS ubiquitin-specific protease; Uspsy gene.
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 Hall, N.M.
AUTHORS Thesis (2001) Department of Pathology, University of Cambridge,
JOURNAL Cambridge, United Kingdom
REFERENCE 2 (bases 1 to 8117)
AUTHORS Hall, N.M.
TITLE Direct Submission
JOURNAL Submitted (02-MAY-2001) Hall N.M., Department of Pathology,
University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP,
UNITED KINGDOM

FEATURES
source Location/Qualifiers
1..8117
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/chromosome="Y"
/map="Xrb"
/sex="Male"
/tissue_type="testis"
1..8117
/gene="Uspsy"
/gene="Uspsy"
419..8089
/gene="Uspsy"
/codon_start=1
/product="putative ubiquitin-specific protease"
/protein_id="CAC38831.1"
/db_xref="GI:14041782"

translation="MTITKSGPVGENSGQSDGQPOPSFOONQISSSDSNETSP
TTPYEQOGAPQHEEDPFPHTDPAKLEMDNRSRWVVPVPEKELEVLLETISI
DLTKGLDVKSEACQRFQDVLTFSKILMDKVAAGKRWVPIINTRHLTLCV
AKLSODMPPFLAIALNPCKFHYNGARCESVSSVQPEDELFAQSPDLSPK
GMVLDLINTFGTLNGFOILHDFRTSGALNTOIAPKPGQCEPFTOTRLTFFI
PIVEVPOILKUTNEBKERTYKADTSMITKPLNLSLRPGQEVYKNTET
RLKTLRLQIISFNGKALNEIKVLSVYTHGSEEWLTYERTWVLCQV
NILSTIVLQDSLHQPVYKIEKILKPVKEKALTLQDINNTMAQAGKEAIVKXVD
LLANLWNSDQVVDIMDLASAKIKILDYSGQDRDQKIQIDCFIEPPTNNKVI
PALKOIKEIGLFEAPONLSQTHSPAFVFRHDLISQLNHALVTVAENLAAYN
SIRLYARHEDVDPQVTLGRSHVQVQERLNLRLKQGLMLCVSQKQIWC
LAENAVYSDBEACFMYSKMGDEPDLHPDINKEFESNVQLDPLSTENGMKCFE
RFPKTVNREGKLMKTKIYMDLDLGLDYLMKVVYQSDNISRAIDLKEIYTS
LGRKQANQVYIHEDFIQSCFDRLKASVDTLCVLDSEKDNIFSCARQAIKVRILTV
LREYISEYSDYHEERMLPMSRAFGKHLSTVAFPGQKEVEDLILSHNATIS
VRRCILNNVAVAHKIELFELGELVASEDRKLEQNLKDKSLITAKETIOTNSM
PSPSSSSSAGPGRSHNNRVNVSPEMKCPGYMSIQRYISFLQVADLGS
MLTVPTLADGARIIMKLMPPDSTTEQIRALCSQDVNIGERLQGSLSLFFGSSASQ
VLVLTVEVYTLMPAGAPLADISDFQYHPLKSGGLPLVLSMLIQNNFLPNTDVTNR
DAVFSALKIATLLTIVGIGHVAIAEACQVADGTDPKTPINQVTHDOAVVLQNALQ

SIRPSSRCMLRNVSHLAQOISGLASRYEDICVIRAIQKIMAGCCSLELVFSPN
EDITETTYMTTSTNSNLEKDEQCEBELLVMTLFCFALIPTAMDSLNKRNKQSVYID
LLVCPSKTRVLAQEOFFLICTRCCKMHRPLFPFTLLFTLLGSAANEKGRSDVYF
TLRLRLTYAVNSNIQVNPVVDLNDLMDKRVADYIKNTGETMAGELPILEGHGV
KELLSFOSPEKRYHIGCKTGANLVKELIDYFIPASXAYLOYMSEGLPIFOALPVC
GSPATINAGFELVALAFGCVRNKKOINVCULTELYIGTPTTCFVAGEMELPVPVQ
RPPGFVGLKNAGATCYMNSVYIQQFMTPSIRNSLAIADISWTDIDDI FKEKQDSE
NNVDREDVFRYPHOEDKPTLSKVEDKEKYNIAVNLQITFGHLLASOLOYYVPRG
FWQGRFLMGEPVNLREOHDALEFNLSLVDSLEDAKALGYPTVSKVLGSGADOKIC
QGCRRYCESESFTLLNVDINRHNOLDLSDIOYVGDLLLEGANAHCCKDKVOTVK
RLIKLPSVLTIOLKREDYMERECALKFNDYFEPPELMEPYTVAGATLKEGDSV
NPOTOLIKONQSESVIPGSTKYRLVGLVHSGONGGHVSYIIORNGKDSKSHMF
KFDGDEYECMDDEEMKNOCCFGEYVGEVFDHMKRMSYRQGRMKNATYLFERM
DITBDEDEITYISELFTFRPHQINSPAIENSVMKONQFPHNQFSLRYFOFIKL
LTCNAVYLSPPAGQDHLLEAEDITMISIOASRFLFTTGFHTKKIIRGPAFSDAL
CILRHSKNVAFWEPYHNVLFVNSNRSFSELYECSPAERIGTFAXLIVFIRAFSDQDS
SPSPFTSPFANPQPSQIYDNLSDSLDLKLVLSLREVSCHGHLQOYFPLFYMA
SLGLAEKTOILKLVNVPATFPMVSLDEGCPVQYAYALSKHSVSLDIRCCSVSR
MOSSINGNPLPFPNPGDNLQPIPIQONVADILFMTTTHKKYIEDCSNEDYVK
LIFCCWENPQFSCVLSLELMQVASHAYELQPYLDLLOIILFEDSOWARIKHAKL
GIPNDQDLFDITIOHSHKQKRAVOCIKMVTLPNSCPVAYOILQGGNDLKNKTWA
MEMWGLDELERKPYSGNPQYTYSNMSPVQSNSTANQYFLEKSHSAKMLTKACDLYPE
EDPDODALDEHVAHQDRTFFVLYSHRSYVQONVVPQPSGPASHHLNPNQKNDKP
OETHSNEISCSLJKQ"

BASE COUNT 2634 a 1481 c 1586 g 2416 t
ORIGIN

Alignment Scores:
Pred. No.: 417 Length: 8117
Score: 46.00 Matches: 9
Percent Similarity: 76.47% Conservative: 4
Best Local Similarity: 52.94% Mismatches: 4
Query Match: 56.10% Indels: 0
Gaps: 0

US-10-081-935-2 (1-18) x MMU307017 (1-8117)

QY 1 MetGlnPheSerHisAlaLeuIleAlaAlaGlyLeuAlaSer 17
Db 2231 CTTGACATAATCATGCTTTAGTACTTGATGACAGAAACCTTGACACT 2281

RESULT 27
LOCUS AX305688 10323 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 439 from Patent WO018188.
ACCESSION AX305688
VERSION AX305688.1 GI:17645121
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 Ishikawa, K., Asai, S., Takahashi, Y., Nagata, T. and Ishii, Y.
AUTHORS Method for examining ischemic conditions
JOURNAL Patent: WO 018188-A 439 22-NOV-2001.
TITLE School Unidical Paris Nihon University (JP)

FEATURES
source Location/Qualifiers
1..10323
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

BASE COUNT 3055 a 1989 c 2201 g 3078 t
ORIGIN

Alignment Scores:
Pred. No.: 532 Length: 10323
Score: 46.00 Matches: 9
Percent Similarity: 76.47% Conservative: 4
Best Local Similarity: 52.94% Mismatches: 4
Query Match: 56.10% Indels: 0
Gaps: 0

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 3, 2003 this sequence version replaced gi:29373197.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H_AA087252A06

FEATURES
source
1. 26013
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/clone="XXFOS-87252A6"

BASE COUNT 6788 a 5982 c 5844 g 7399 t
ORIGIN

Alignment Scores:
Pred. No.: 1.36e+03 Length: 26013
Score: 46.00 Matches: 11
Percent Similarity: 78.57% Conservative: 0
Best Local Similarity: 78.57% Mismatches: 3
Query Match: 56.10% Indels: 0
DB: 9 Gaps: 0

US-10-081-935-2 (1-18) x AC142230 (1-26013)

OY 2 GlnPhSeRHisAlaLeuIleAlaLeuValAlaAlaGlyLeu 15

Db 5372 CAGTTCTCCACGCCCTTCCAGCCTTCTTCAGCAGCGCCCTT 5413

RESULT 30 AC138990 125439 bp DNA linear PRI 27-MAR-2003

LOCUS Homo sapiens chromosome 5 clone RP13-708P15, complete sequence.
AC138990
AC138990.2 GI:29294040
VERSION
KEYWORDS
SOURCE
ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 125439)
DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission
Unpublished

2 (bases 1 to 125439)
DOE Joint Genome Institute.

Submitted (22-JAN-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 125439)

DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission

Submitted (27-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Mar 27, 2003 this sequence version replaced gi:27819484.

Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu

Quality: Phrap Quality >40 100% of Sequence;
Estimated Total Number of Errors is 0.

NOTE: This sequence is not the entire sequence of the clone
(entire sequence is 153.5kb). It is clipped at the overlap with
AC114316. The number of basepairs overlapped is 118195.

FEATURES
source
1. 125439
Location/Qualifiers

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP13-708P15"

Alignment Scores:
Pred. No.: 6.73e+03 Length: 125439
Score: 46.00 Matches: 9
Percent Similarity: 81.25% Conservative: 4
Best Local Similarity: 56.25% Mismatches: 3
Query Match: 56.10% Indels: 0
DB: 9 Gaps: 0

US-10-081-935-2 (1-18) x AC138990 (1-125439)

OY 2 GlnPhSeRHisAlaLeuIleAlaLeuValAlaAlaGlyLeu 17

Db 43208 AATACCATCATGCGCTATTAGCAAAAGTGCGAGTTGCTTATCTCA 43255

RESULT 31 AC073440 133445 bp DNA linear HTG 08-JAN-2003
LOCUS Homo sapiens chromosome 12 clone RP11-100E14, WORKING DRAFT
DEFINITION
SEQUENCE, 4 unordered pieces.
AC073440
AC073440.12 GI:20335621
VERSION
KEYWORDS
HTG: HTGS PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 133445)

REFERENCE
AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barberia,J., Benton,J., Bimge,K., Binkenburg,K., Bonini,D.,
Bouck,J., Bowe,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Cairon,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyte,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroli,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.U., Draper,H., Dugan-Rocha,S., Durkin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsf.,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlovic,J., Kurehi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,U., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W., Louieged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,M., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogun,M., Okwono,G.,
Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., RojuboKan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshari,N., Stason,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Sytek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Taneey,J., Taylor,C., Taylor,T., Tellod,B., Thomas,N., Thomas,S.,
Uemami,K., Vasquez,L., Vera,V., Villalton,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,

Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 133445)
Mortley, K. C.
Direct Submission
Submitted (18-JUN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 133445)
Mortley, K. C.
Direct Submission
Submitted (08-JAN-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:16117941.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HBCF
Center clone name: RP11-100E14
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-terminator Big Dye; 81% of reads
Assembly program: Phrap; version 0.990325
Consensus quality: 128546 bases at least Q40
Consensus quality: 130561 bases at least Q30
Consensus quality: 131661 bases at least Q20
Estimated insert size: 134201; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1 2013: contig of 2013 bp in length
2014 2113: gap of unknown length
2114 31789: contig of 29676 bp in length
31790 31889: gap of unknown length
31890 79584: contig of 47695 bp in length
79585 79684: gap of unknown length
79685 133445: contig of 53761 bp in length.
Location/Qualifiers
1.133445
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-100E14"
BASE COUNT 41186 a 25768 c 25551 g 40640 t 300 others
ORIGIN
Alignment Scores:
Pred. No.: 7.16e+03 Length: 133445
Score: 46.00 Matches: 7
Percent Similarity: 93.75% Conservative: 8
Best Local Similarity: 43.75% Mismatches: 1
Query Match: 56.10% Indels: 0
Gaps: 0
US-10-081-935-2 (1-18) x AC073440 (1-133445)

Cy 1 MetClnpheserHisAlaLeuIleAlaLeuValAlaIaGlyLeuAla 16
Db 117092 TTGAGCTTTACTACTCTCTCTCCCTTCGTGAGGCTGAAATGCCA 117045
RESULT 32
AC138980 147008 bp DNA linear HTG 22-JAN-2003
AC138980
LOCUS
DEFINITION
Homo sapiens chromosome 5 clone RP13-169F15, WORKING DRAFT
SEQUENCE, 2 unordered pieces.
AC138980 147008 bp DNA linear HTG 22-JAN-2003
AC138980
AC138980.1 GI:27819474
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
DOE Joint Genome Institute.
Direct Submission
Submitted (22-JAN-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
----- Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: <http://www.jgi.doe.gov>
----- Project Information
Center Project Name: 3069541
Center clone name: RPCT-13_169F15
----- Summary Statistics
Consensus quality: 146757 bases at least Q40
Consensus quality: 146840 bases at least Q30
Consensus quality: 146855 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 146908; sum-of-contigs estimation
Quality coverage: 15.57 in Q20 bases; agarose-fp estimation
Quality coverage: 18.54 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 28773: contig of 28773 bp in length
28774 28874: gap of unknown length
28874 147008: contig of 118135 bp in length.
Location/Qualifiers
1.147008
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP13-169F15"
/clone_1fb="RPCT human BAC library 13"
BASE COUNT 48290 a 26527 c 25955 g 46136 t 100 others
ORIGIN
Alignment Scores:
Pred. No.: 7.9e+03 Length: 147008
Score: 46.00 Matches: 9
Percent Similarity: 81.25% Conservative: 4
Best Local Similarity: 56.25% Mismatches: 3
Query Match: 56.10% Indels: 0
Gaps: 0

```

US-10-081-935-2 (1-18) x AC138980 (1-147008)
QY      2 GlnPhseSerHisAlaLeuIleAlaLeuValAlaIleGlyLeuAlaSer 17
       : ::::: ||||| ::::: ||||| ||||| ::::: |||||
Db      72017 AAATACCATCATCGCCCTATTAGCAAAATGCGACGTTGGTATCCTCA 72064

RESULT 33
AC138981      147177 bp      DNA      linear      HTG 22-JAN-2003
LOCUS      AC138981
DEFINITION      Homo sapiens chromosome 5 clone RP13-193110, WORKING DRAFT
ACCESSION      AC138981
VERSION      AC138981.1 GI:27819475
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 147177)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 147177)
DOE Joint Genome Institute.
Direct Submission
Submitted (22-JAN-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

-----
Project Information
Center Project Name: 3078824
Center clone name: RPC1-13_193110
-----
Summary Statistics
Consensus quality: 146200 bases at least Q40
Consensus quality: 146424 bases at least Q30
Consensus quality: 146579 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 146777; sum-of-ctnigs estimation
Quality coverage: 5.31 in Q20 bases; agarose-fp estimation
Quality coverage: 6.33 in Q20 bases; sum-of-ctnigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 5 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2663: ctnig of 2662 bp in length
2663      2762: gap of unknown length
2763      12052: ctnig of 9290 bp in length
12053      12152: gap of unknown length
12153      28841: ctnig of 16789 bp in length
28842      29041: gap of unknown length
29042      86682: ctnig of 57641 bp in length
86683      86782: gap of unknown length
86783      147177: ctnig of 60395 bp in length.
Location/Qualifiers
1..147177
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP13-193110"
/clone_lib="RPC1 human BAC library 13"
BASE COUNT      48517 a 26217 c 26202 g 45840 t 401 others
ORIGIN
Alignment Scores:

```

```

Pred. No.:      7_9le+03      Length:      147177
Score:          46.00        Matches:       9
Percent Similarity: 81.25%   Conservative: 4
Best Local Similarity: 56.25% Mismatches:    3
Query Match:     56.10%     Indels:         0
DB:              2           Gaps:          0

US-10-081-935-2 (1-18) x AC138981 (1-147177)

OY      2 GlNPheserHisalaleUjaIealaLeValaIAlaGlyLeuaJaseR 17
Db      72151 AAMTACCATCATGCCCTATTAGCAAAAGTGCGACTTCATCCTCA 72198
                ::::: ||||| |::: ||||| |::: ||||| :|
RESULT 34 AP001493/c      161624 bp DNA linear HTG 30-MAY-2000
LOCUS      AP001493/c      DRAFT SEQUENCE, 23 unordered pieces.
DEFINITION
ACCESSION  AP001493.2 GI:8117351
VERSION     AP001493.2
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eureleostomi;
REFERENCE   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Published Only in DataBase (2000)
JOURNAL     2 (bases 1 to 161624)
HATTORI,M., ISHII,K., TOYODA,A., TAYLOR,T.D., HONG-SEOG,P., FUJIYAMA,A.,
YADA,T., TOTOKI,Y., WATANABE,H. AND SAKAKI,Y.
Fujiyama,A., Yada,T., Totoki.Y., Watanabe.H. and Sakaki.Y.
Direct Submission
Submitted (17-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Science Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced gi:7288187.

COMMENT----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information -----
Center project name: Humraltit8
Center clone name: RP11-779H19
----- Summary Statistics -----
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amstrham; 100% of reads
Assembly program: Phrap; version O.990329
Consensus quality: 150339 bases at least Q40
Consensus quality: 155891 bases at least Q30
Consensus quality: 158205 bases at least Q20
Insert size: 159424; sum-of-contigs
Quality coverage: 5.50x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
23 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1      32754 contig of 32754 bp in length
32855  48719 contig of 15865 bp in length
48820  64088 contig of 15269 bp in length
64189  76256 contig of 12068 bp in length
76357  86295 contig of 9939 bp in length
86396  94552 contig of 8157 bp in length
94653  102595 contig of 7943 bp in length
```

102696 110667 contig of 7972 bp in length
110768 116555 contig of 5788 bp in length
116656 122099 contig of 5444 bp in length
122000 125461 contig of 3262 bp in length
125562 131077 contig of 5516 bp in length
131178 135924 contig of 4747 bp in length
136025 138309 contig of 2285 bp in length
138410 140832 contig of 2423 bp in length
140933 144613 contig of 3681 bp in length
144714 147380 contig of 2667 bp in length
147481 150820 contig of 3340 bp in length
150921 154114 contig of 3194 bp in length
154215 156288 contig of 2074 bp in length
156389 158482 contig of 2094 bp in length
158583 160442 contig of 1860 bp in length
160543 161624 contig of 1082 bp in length

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 32754: contig of 32754 bp in length
32755 32854: gap of 100 bp
32855 48719: contig of 15865 bp in length
48720 48819: gap of 100 bp
48820 64088: contig of 15269 bp in length
64089 64188: gap of 100 bp
64189 76256: contig of 12068 bp in length
76257 86295: contig of 9939 bp in length
86296 86395: gap of 100 bp
86396 94552: contig of 8157 bp in length
94553 94652: gap of 100 bp
94653 102595: contig of 7943 bp in length
102596 102695: gap of 100 bp
102696 110667: contig of 7972 bp in length
110668 110768: gap of 100 bp
110769 116555: contig of 5788 bp in length
116556 116656: gap of 100 bp
116657 122099: contig of 5444 bp in length
122100 122199: gap of 100 bp
122200 125461: contig of 3262 bp in length
125462 125561: gap of 100 bp
125562 131077: contig of 5516 bp in length
131078 131177: gap of 100 bp
131178 135924: contig of 4747 bp in length
135925 136024: gap of 100 bp
136025 138309: contig of 2285 bp in length
138310 138409: gap of 100 bp
138410 140832: contig of 2423 bp in length
140833 140933: gap of 100 bp
140934 144613: contig of 3681 bp in length
144614 144713: gap of 100 bp
144714 147380: contig of 2667 bp in length
147381 150820: contig of 3340 bp in length
150821 150921: gap of 100 bp
150922 154114: contig of 3194 bp in length
154115 154215: gap of 100 bp
154216 156288: contig of 2074 bp in length
156289 156389: gap of 100 bp
156390 158482: contig of 2094 bp in length
158483 158583: gap of 100 bp
158584 160442: contig of 1860 bp in length
160443 160542: gap of 100 bp
160543 161624: contig of 1082 bp in length.

FEATURES
Source
1. .161624
/organism="Homo sapiens"

/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-779H19"
1. .32754
/note="assembly_fragment"
32855. .48719
/note="assembly_fragment clone_end:T7 vector_side:left"
48820. .64088
/note="assembly_fragment"
64189. .76256
/note="assembly_fragment"
76357. .86295
/note="assembly_fragment"
86396. .94552
/note="assembly_fragment"
94653. .102595
/note="assembly_fragment"
102696. .110667
/note="assembly_fragment"
110768. .116555
/note="assembly_fragment"
116556. .122099
/note="assembly_fragment"
122000. .125461
/note="assembly_fragment"
125562. .131077
/note="assembly_fragment"
131178. .135924
/note="assembly_fragment"
136025. .138309
/note="assembly_fragment clone_end:SP6 vector_side:left"
138410. .140832
/note="assembly_fragment"
140933. .144613
/note="assembly_fragment"
144714. .147380
/note="assembly_fragment"
147481. .150820
/note="assembly_fragment"
150921. .154114
/note="assembly_fragment"
154215. .156288
/note="assembly_fragment"
156389. .158482
/note="assembly_fragment"
158583. .160442
/note="assembly_fragment"
160543. .161624
/note="assembly_fragment"

BASE COUNT 49551 a 30115 c 30590 g 49168 t 2200 others
ORIGIN

Alignment Scores:
Pred. No.: 8.7e+03 Length: 161624
Score: 46.00 Matches: 7
Percent Similarity: 93.75% Conservative: 8
Best Local Similarity: 43.75% Mismatches: 1
Query Match: 56.10% Indels: 0
DB: 2 Gaps: 0

US-10-081-935-2 (1-18) x AP001493 (1-161624)

Qy 1 MercInPhSeSerHsAlaLeuIleAlaLeuValAlaIaGlyLeuAla 16
Db 120229 TTGAGGTTTACTACTCTCTTCCTTCCTGAGGCTGGAATGCCA 120182

RESULT 35
AL391259 163520 bp DNA linear PRI 06-DEC-2001
LOCUS Human DNA sequence from clone RP11-469E19 on chromosome Xp11.4-21.2
DEFINITION Contains a pseudogene similar to chloride intracellular channel 1

protein, the 5' end of the USP9X gene for ubiquitin specific protease 9 (Drosophila fat facets related) and a Cpg island, complete sequence.

ACCESSION AL391259
VERSION AL391259.15 GI:11322864
KEYWORDS HTG; chloride; Cpg island; intracellular; protease; ubiquitin; USP9X.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 163520)
TITLE Whitehead, S.
AUTHORS
JOURNAL
COMMENT
Submitted (06-DEC-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humbrey@sanger.ac.uk
On Nov 23, 2000 this sequence version replaced gi11322140.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-469E19 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-469E19 it may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-469E19 is at 1 in this sequence.
The true left end of clone RP5-1051C18 is at 185 in this sequence.
The true left end of clone RP5-1172N10 is at 163421 in this sequence.

FEATURES
source
1..163520
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="p11.4-21.2"
/clone="RP11-469E19"
/clone_11b="RPCI-11.2"
408..615
repeat_region
/note="AluIo repeat: matches 87. .286 of consensus"
1462..1669
repeat_region
/note="Wt1D repeat: matches 217. .461 of consensus"
1726..1897
repeat_region
/note="FRAM repeat: matches 0. .173 of consensus"
1923..2107
repeat_region
/note="Wt1D repeat: matches 3. .184 of consensus"
2147..2448
repeat_region
/note="AluSc repeat: matches 1. .302 of consensus"
2490..2647
repeat_region
/note="AluSc repeat: matches 139. .296 of consensus"
2678..2970
repeat_region

misc_feature
complement(12970..3296)
/note="match: GSS: Em:AQ279697"
2983..3228
repeat_region
/note="AluIb repeat: matches 24. .293 of consensus"
complement(3221..3328)
/note="match: GSS: Em:AQ278728"
4145..4279
repeat_region
/note="FLAM_C repeat: matches 12. .143 of consensus"
4693..4907
repeat_region
/note="AluIb repeat: matches 85. .309 of consensus"
4918..5199
repeat_region
/note="AluSc repeat: matches 1. .282 of consensus"
5200..5275
repeat_region
/note="38 copies 2 mer ta 97% conserved"
5570..5872
repeat_region
/note="AluSc repeat: matches 1. .303 of consensus"
6323..6576
repeat_region
/note="AluIo repeat: matches 40. .281 of consensus"
6627..6702
repeat_region
/note="LTRJ3 repeat: matches 2. .84 of consensus"
6761..6931
repeat_region
/note="WIR repeat: matches 14. .191 of consensus"
7009..7140
repeat_region
/note="AluIo repeat: matches 1. .136 of consensus"
7141..7435
repeat_region
/note="AluSP repeat: matches 1. .296 of consensus"
7436..7625
repeat_region
/note="AluIo repeat: matches 136. .312 of consensus"
7626..7733
repeat_region
/note="54 copies 2 mer aa 77% conserved"
8520..8657
repeat_region
/note="WIR repeat: matches 3. .147 of consensus"
8850..8977
repeat_region
/note="AluX repeat: matches 9. .137 of consensus"
8978..9298
repeat_region
/note="AluSg repeat: matches 1. .308 of consensus"
9299..9451
repeat_region
/note="AluX repeat: matches 137. .286 of consensus"
9567..9919
repeat_region
/note="L1M5 repeat: matches 7156. .7496 of consensus"
9924..10086
repeat_region
/note="Char1el1 repeat: matches 2593. .2756 of consensus"
10089..10233
repeat_region
/note="FRAM repeat: matches 2. .146 of consensus"
10256..10557
repeat_region
/note="AluX repeat: matches 3. .304 of consensus"
10568..10946
repeat_region
/note="Char1el1 repeat: matches 2161. .2574 of consensus"
10945..11034
repeat_region
/note="Char1el2 repeat: matches 1287. .1379 of consensus"
11039..11322
repeat_region
/note="AluSc repeat: matches 2. .302 of consensus"
11333..11367
repeat_region
/note="U2 repeat: matches 1. .35 of consensus"
11361..11519
repeat_region
/note="FRAM repeat: matches 2. .146 of consensus"
11630..12322
repeat_region
/note="Char1el2 repeat: matches 381. .1104 of consensus"
12345..12460
repeat_region
/note="FLAM_A repeat: matches 1. .115 of consensus"
12480..12804
repeat_region
/note="Char1el2 repeat: matches 33. .367 of consensus"
12896..13033
repeat_region
/note="L1M5/D repeat: matches 5662. .5791 of consensus"
13077..13178
repeat_region
/note="FLAM_A repeat: matches 27. .109 of consensus"
13184..13491
repeat_region
/note="L1M5/D repeat: matches 5323. .5613 of consensus"
13500..13636
repeat_region
/note="LTR41 repeat: matches 13. .148 of consensus"
13637..13940
repeat_region
/note="AluX repeat: matches 1. .305 of consensus"

```

repeat_region      13941..14089
                    /note="LTR41 repeat: matches 148..301 of consensus"
repeat_region      14112..14144
                    /note="LTR41 repeat: matches 418..433 of consensus"
repeat_region      14145..14452
                    /note="ALUdo repeat: matches 1..305 of consensus"
repeat_region      14852..15028
                    /note="L1MD1 repeat: matches 5705..5884 of consensus"
repeat_region      15127..15182
                    /note="LTR23 repeat: matches 204..436 of consensus"
repeat_region      15383..15480
                    /note="BUR1 repeat: matches 6167..6271 of consensus"
repeat_region      15482..15563
                    /note="LTR23 repeat: matches 36..109 of consensus"
repeat_region      15595..15668
                    /note="L1P4 repeat: matches 5258..5451 of consensus"
repeat_region      15980..16075
                    /note="MER5A repeat: matches 13..112 of consensus"
repeat_region      16796..17092
                    /note="ALUSP repeat: matches 1..296 of consensus"
repeat_region      17479..17743
                    /note="ALUSx repeat: matches 1..266 of consensus"
repeat_region      17743..17987
                    /note="ALUy repeat: matches 46..290 of consensus"
repeat_region      18468..18553
                    /note="MLT1H repeat: matches 93..172 of consensus"
repeat_region      18606..18832
                    /note="ALUSg/x repeat: matches 70..296 of consensus"
repeat_region      18833..18998
                    /note="ALUSx repeat: matches 1..166 of consensus"
repeat_region      19100..19227
                    /note="L1PA11 repeat: matches 5718..5849 of consensus"
repeat_region      19346..19646
                    /note="ALUSg repeat: matches 1..301 of consensus"
repeat_region      19647..19730
                    /note="42 copies 2 mer ta 65% conserved"
repeat_region      19835..20139
                    /note="ALUSx repeat: matches 3..297 of consensus"
repeat_region      20954..21192
                    /note="ALUub repeat: matches 77..311 of consensus"
repeat_region      21361..21479
                    /note="ALUSg/x repeat: matches 1..120 of consensus"
repeat_region      21492..21791
                    /note="ALUSg repeat: matches 1..293 of consensus"
repeat_region      21922..22076
                    /note="MER5A repeat: matches 26..189 of consensus"
repeat_region      22091..22191
                    /note="MIR repeat: matches 33..134 of consensus"
repeat_region      22249..22928
                    /note="MER51B repeat: matches 1..617 of consensus"
repeat_region      23550..23861

Alignment Scores:
Pred. No.:      8.8e+03      Length:      163520
Score:          46..00      Matches:      9
Percent Similarity: 76..47%      Conservative: 4
Best Local Similarity: 52..94%      Mismatches: 4
Query Match:     56..10%      Indels:      0
DB:              9          Gaps:      0

US-10-081-935-2 (1-18) x ALJ391259 (1-163520)
Oy      1 MetGlnPheserHisAlaLeuIleAlaLeuValAlaIaGlyLeuAlaSer 17
Db      142090 CTTCAACACATCATGCCCTAGTACTTGTGACACAAAACCTTGCACT 142140

RESULT 36
ALJ356864/c      166876 bp      DNA      linear      HTG 13-JUN-2001
LOCUS
DEFINITION      Homo sapiens chromosome 1 clone RP5-1051C18, *** SEQUENCING IN
PROGRESS ***, 26 unordered pieces.
ACCESSION      ALJ356864
VERSION      ALJ356864.6      GI:9926631

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

HTG: HTGS PHASE1; HTGS_CANCELLED.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
Plumb, B.
Direct Submission
Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequests@sanger.ac.uk
On Aug 25, 2000 this sequence version replaced gi:9214009.

Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

Project Information
Center project name: dj1051C18

Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 151580 bases at least Q40
Consensus quality: 156291 bases at least Q30
Consensus quality: 159609 bases at least Q20
Insert size: 164376; sum-of-contigs
Insert size: 184764; 6.6% error; agarose-fp
Quality coverage: 2.91x in Q20 bases; sum-of-contigs Quality
coverage: 2.72x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
7483: contig of 7483 bp in length
7484 7583: gap of 100 bp
7584 1184: contig of 3601 bp in length
1185 1184: gap of 100 bp
1185 18742: contig of 7458 bp in length
18743 18842: gap of 100 bp
18843 22744: contig of 3902 bp in length
22745 22844: gap of 100 bp
22845 30191: contig of 7347 bp in length
30192 30291: gap of 100 bp
30292 36951: contig of 6660 bp in length
36952 37051: gap of 100 bp
37052 41101: contig of 4050 bp in length
41102 41201: gap of 100 bp
41202 44151: contig of 2950 bp in length
44152 44251: gap of 100 bp
44252 48108: contig of 3857 bp in length
48109 48208: gap of 100 bp
48209 58414: contig of 10206 bp in length
58415 58514: gap of 100 bp
58515 77249: contig of 18735 bp in length
77250 77349: gap of 100 bp
77350 82656: contig of 5307 bp in length
82657 82756: gap of 100 bp
82757 85452: contig of 2696 bp in length
85453 85552: gap of 100 bp
85553 90378: contig of 4826 bp in length
90379 90478: gap of 100 bp
90479 95479: contig of 4995 bp in length
95479 95573: gap of 100 bp
95574 98653: contig of 2980 bp in length
98654 98654: gap of 100 bp
98655 105535: contig of 6882 bp in length
105536 105536: gap of 100 bp

* 105636 114741: contig of 9106 bp in length
* 114742 114841: gap of 100 bp
* 114842 118544: contig of 3703 bp in length
* 118545 118644: gap of 100 bp
* 118645 130337: contig of 11693 bp in length
* 130338 130437: gap of 100 bp
* 130438 144196: contig of 13759 bp in length
* 144197 144296: gap of 100 bp
* 144297 149313: contig of 5017 bp in length
* 149314 149314: gap of 100 bp
* 149314 152428: contig of 3015 bp in length
* 152429 152528: gap of 100 bp
* 154910 154910: contig of 2382 bp in length
* 154911 155010: gap of 100 bp
* 155011 163649: contig of 8639 bp in length
* 163650 163749: gap of 100 bp
* 163750 166876: contig of 3127 bp in length.
FEATURES
source
1..166876
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP5-1051C18"
/clone_11b="RPC1-5"
1..7483
/note="assembly_fragment:01007
clone_end:77
vector_side:left"
7584..11184
/note="assembly_fragment:00504
fragment_chain:1"
11285..18742
/note="assembly_fragment:00626
fragment_chain:1"
18843..22744
/note="assembly_fragment:00511
fragment_chain:2"
22845..30191
/note="assembly_fragment:01426
fragment_chain:2"
30292..36951
/note="assembly_fragment:00759
fragment_chain:3"
37052..41101
/note="assembly_fragment:01817
fragment_chain:3"
41202..44151
/note="assembly_fragment:00825
fragment_chain:4"
44252..48108
/note="assembly_fragment:01678
fragment_chain:4"
48209..58414
/note="assembly_fragment:01296
fragment_chain:5"
58515..77249
/note="assembly_fragment:00448
fragment_chain:5"
77350..82656
/note="assembly_fragment:01760
fragment_chain:6"
82757..85452
/note="assembly_fragment:00800
fragment_chain:6"
85553..90378
/note="assembly_fragment:00196"
90479..95473
/note="assembly_fragment:00312"
95574..98553
/note="assembly_fragment:00433"
98654..105535
/note="assembly_fragment:00502"

misc_feature 105636..114741
/note="assembly_fragment:00551"
misc_feature 114842..118544
/note="assembly_fragment:00557"
misc_feature 118645..130337
/note="assembly_fragment:00701"
misc_feature 130438..144196
/note="assembly_fragment:00873"
misc_feature 144297..149313
/note="assembly_fragment:01103"
misc_feature 149414..152428
/note="assembly_fragment:01347"
misc_feature 152529..154910
/note="assembly_fragment:01545"
misc_feature 155011..163649
/note="assembly_fragment:01263
fragment_chain:7"
163750..166876
/note="assembly_fragment:01448
fragment_chain:7
clone_end:SP6
vector_side:right"
BASE COUNT 51122 a 33170 c 31820 g 48259 t 2505 others
ORIGIN
Alignment Scores:
Pred. No.: 8.99e+03 Length: 166876
Score: 46.00 Matches: 9
Percent Similarity: 76.47% Conservative: 4
Best Local Similarity: 52.94% Mismatches: 4
Query Match: 56.10% Indels: 0
DB: 2 Gaps: 0
US-10-081-935-2 (1-18) x AL356864 (1-166876)
Cy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaGlyLeuAlaSer 17
Db 68948 CTTCAACACATCATGCCCTAGTACTTGTGTACAGAAACCTTGCACT 68898
RESULT 37
AC118440/c 169693 bp DNA linear HTG_19-NOV-2002
LOCUS Rattus norvegicus clone CH230-255F21, WORKING DRAFT SEQUENCE, 6
DEFINITION Rattus norvegicus clone CH230-255F21, WORKING DRAFT SEQUENCE, 6
ACCESSION AC118440
VERSION AC118440.4 GI:25099817
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 169693)
REFERENCE
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Angiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biewald,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabiri,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Huylk,S., Hume,J., Idlebird,D., Jackson,A.,

[illegible]

```

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 71663: contig of 71663 bp in length
*
* 71664 71763: gap of unknown length
*
* 71764 162672: contig of 90909 bp in length
*
* 162673 162772: gap of unknown length
*
* 162773 164075: contig of 1303 bp in length
*
* 164076 164175: gap of unknown length
*
* 164176 165368: contig of 1193 bp in length
*
* 165369 165468: gap of unknown length
*
* 165469 167978: contig of 2510 bp in length
*
* 167979 168078: gap of unknown length
*
* 168079 169693: contig of 1615 bp in length.
*
* Location/Qualifiers
*
* 1. 169693
*
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-255F21"
* 66802..68280
* /note="wgs contig"
*
* 69386..71663
* /note="wgs contig"
*
* complement(88241..88650)
* /note="clone boundary
* clone_end:T7
* site:
*
* end_sequence:B2115992"

```


JOURNAL

Submitted (24-MAR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924)

COMMENT

On May 30, 2000 this sequence version replaced gi:7340850.

----- Genome Center

Center: RIKEN Genomic Sciences Center (GSC)

Center code: RIKEN

Web site: http://hgp.gsc.riken.go.jp/

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDrafl18

Center clone name: RP11-83712

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 162012 bases at least Q40

Consensus quality: 165830 bases at least Q30

Insert size: 168670; sum-of-contigs

Quality coverage: 5.46x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 30841 contig of 30841 bp in length
30942 49995 contig of 19054 bp in length
50096 70667 contig of 20572 bp in length
70768 89290 contig of 18523 bp in length
89391 105447 contig of 16057 bp in length
105548 120373 contig of 14826 bp in length
120474 133152 contig of 12679 bp in length
133253 143203 contig of 9951 bp in length
143304 147902 contig of 4599 bp in length
148003 152506 contig of 4504 bp in length
152607 155699 contig of 3093 bp in length
155800 158705 contig of 2906 bp in length
158806 162039 contig of 3223 bp in length
162140 164114 contig of 1975 bp in length
164215 166502 contig of 2288 bp in length
166603 169081 contig of 2479 bp in length
169182 170270 contig of 1089 bp in length

```

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 30841: contig of 30841 bp in length
30842 30941: gap of 100 bp
49995 49995: contig of 19054 bp in length
50096 70667: gap of 100 bp
70667 70667: contig of 20572 bp in length
70668 70668: gap of 100 bp
89290 89290: contig of 18523 bp in length
89291 89291: gap of 100 bp
105447 105447: contig of 16057 bp in length
105548 105548: gap of 100 bp
120373 120373: contig of 14826 bp in length
120474 120474: gap of 100 bp
133152 133152: contig of 12679 bp in length
133253 133253: gap of 100 bp
143203: contig of 9951 bp in length

```

```

* 143204 143303: gap of 100 bp
* 143303 147902: contig of 4599 bp in length
* 147902 148002: gap of 100 bp
* 148002 148003: contig of 4504 bp in length
* 148003 152506: contig of 4504 bp in length
* 152506 152607: gap of 100 bp
* 152607 155699: contig of 3093 bp in length
* 155699 155700: gap of 100 bp
* 155700 155799: gap of 100 bp
* 155799 158705: contig of 2906 bp in length
* 158705 158806: gap of 100 bp
* 158806 162039: contig of 3234 bp in length
* 162039 162140: gap of 100 bp
* 162140 164114: contig of 1975 bp in length
* 164114 164215: gap of 100 bp
* 164215 166502: contig of 2288 bp in length
* 166502 166603: gap of 100 bp
* 166603 169081: contig of 2479 bp in length
* 169081 169182: gap of 100 bp
* 169182 170270: contig of 1089 bp in length.

```

FEATURES

source

```

1..170270
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-83712"
1..30841
/note="assembly_fragment"
30942..49995
/note="assembly_fragment"
50096..70667
/note="assembly_fragment"
70768..89290
/note="assembly_fragment"
89391..105447
/note="assembly_fragment"
105448..120373
/note="assembly_fragment"
120374..133152
/note="assembly_fragment"
133153..143203
/note="assembly_fragment"
143204..147902
/note="assembly_fragment"
147903..152506
/note="assembly_fragment"
152507..155699
/note="assembly_fragment"
155699..158705
/note="assembly_fragment"
158706..162039
/note="assembly_fragment"
162040..164114
/note="assembly_fragment"
164115..166502
/note="assembly_fragment"
166503..169081
/note="assembly_fragment"
169082..170270
/note="assembly_fragment"

```

BASE COUNT 51510 a 32818 c 32593 g 51749 t 1600 others

ORIGIN

Alignment Scores:

Pred. No.:	9.17e+03	Length:	170270
Score:	46.00	Matches:	7
Percent Similarity:	93.75%	Conservative:	8
Best Local Similarity:	43.75%	Mismatches:	1
Query Match:	56.10%	Indels:	0
DB:	2	Gaps:	0

US-10-081-935-2 (1-18) x AP001548 (1-170270)

QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAla 16
 Db 58369 TTGAGGTTTACTCACTCTCTCTTCCCTTCGAGCTCGAATGCA 58322

RESULT 39
 AC090405 170413 bp DNA linear HTG 27-MAY-2001
 LOCUS AC090405
 DEFINITION Homo sapiens chromosome 18 clone RP11-779H19 map 18, WORKING DRAFT
 AC090405
 AC090405.2 GI:14210571
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 170413)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barnes,N., Baxten,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gargana,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagedorn,B., Heathford,A., Horton,L., Hulme,W., Hultev,I., Johnson,R., Jones,C., Karatas,A., Labrecque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhhang,P., Pierre,N., Polara,V., Raymond,C., Retta,R., Riebeck,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testave,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vasilev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 27, 2001 this sequence version replaced gi:12958041.
 All repeats were identified using RepeatMasker:
 Smit,A.P.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L12714
 Center clone name: 779.H.19

----- Summary Statistics
 Sequencing vector: Plasmid, n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 165803 bases at least Q40
 Consensus quality: 167767 bases at least Q30
 Consensus quality: 168509 bases at least Q20
 Insert size: 157000; agarose-fp
 Insert size: 169313; sum-of-coverage
 Quality coverage: 8.9 in Q20 bases; agarose-fp
 Quality coverage: 8.3 in Q20 bases; sum-of-coverage

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 27912: contig of 27912 bp in length
 * 27913 28012: gap of 100 bp
 * 28013 28759: contig of 747 bp in length
 * 28760 28859: gap of 100 bp
 * 28860 29596: contig of 737 bp in length
 * 29597 29696: gap of 100 bp
 * 29697 30720: contig of 1024 bp in length
 * 30721 30820: gap of 100 bp
 * 30821 31522: contig of 702 bp in length
 * 31523 31623: gap of 100 bp
 * 31623 32368: contig of 746 bp in length
 * 32369 32468: gap of 100 bp
 * 32469 33330: contig of 862 bp in length
 * 33331 33430: gap of 100 bp
 * 33431 34637: contig of 1207 bp in length
 * 34638 34737: gap of 100 bp
 * 34738 38948: contig of 4211 bp in length
 * 38949 39048: gap of 100 bp
 * 39049 43570: contig of 4522 bp in length
 * 43571 43670: gap of 100 bp
 * 43671 162170: contig of 118500 bp in length
 * 162171 162270: gap of 100 bp
 * 162271 170413: contig of 8143 bp in length.

FEATURES
 source
 1. 170413
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone="RP11-779H19"
 /clone_1lb="RPCL-11 Human Male BAC"
 1. 27912
 /note="assembly_fragment"
 clone_end:Spe
 vector_side:left
 28013..28759
 /note="assembly_fragment"
 28860..29596
 /note="assembly_fragment"
 29697..30720
 /note="assembly_fragment"
 30821..31522
 /note="assembly_fragment"
 31623..32368
 /note="assembly_fragment"
 32469..33330
 /note="assembly_fragment"
 33431..34637
 /note="assembly_fragment"
 34738..38948
 /note="assembly_fragment"
 39049..43570
 /note="assembly_fragment"
 43671..162170
 /note="assembly_fragment"
 162271..170413
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right

BASE COUNT 50853 a 32105 c 32644 g 53709 t 1102 others
 ORIGIN

Alignment Scores:
 Pred. No.: 9,18e+03 Length: 170413
 Score: 46.00 Matches: 7
 Percent Similarity: 93.75% Conservative: 8
 Best Local Similarity: 43.75% Mismatches: 1
 Query Match: 56.10% Indels: 0

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 6, 2003, 00:08:38 / Search time 2382 Seconds

(without alignments)
3331.850 Million cell updates/sec

Title: US-10-081-935-4
Perfect score: 1023

Sequence: 1 MOPSHALIALVAGLASAQL.....NVRASVGGIAALLGLAAYL 194

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Delop 6.0 , Deloxt 7.0
Delop 6.0 , Deloxt 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ p2n.model -DEV=xlh
-Q/cg2_1/USFTO.spool/US10081935/runat_04122003_134829_10475/app-query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=-1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pc -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=pro -NORM=exc -HEASize=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081935 @CGN 1.1.2372 @runat_04122003_134829_10475 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELOXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: gb_vl.*
16: em_ba.*
17: em_fun.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pal.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rod.*
36: em_hcg_mam.*
37: em_hcg_vtr.*
38: em_by.*
39: em_higo_hum.*
40: em_higo_mus.*
41: em_higo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB	ID	Description
1	1023	100.0	1201	8	CIU39835	U39835 Coccidioid
2	1023	100.0	1234	8	CIU32518	U32518 Coccidioid
3	970	94.8	1435	8	CIU51200	US1200 Coccidioid
4	970	94.8	3821	8	AF013256	AF013256 Coccidiol
5	206.5	20.2	97478	2	AC105929	AC105929 Magnapor
6	186.5	18.2	2960	3	AY057052	AY057052 Mamestra
7	185.5	18.1	1470	6	AR140164	AR140164 Sequence
8	184	18.0	149696	14	AF19782	AF19782 Callitric
9	182.5	17.8	3655	3	DDISGSP	M3861 Dictyosteli
10	175	17.1	2455	3	TNAF000605	AF000605 Trichoplu
11	175	17.1	2455	6	AR129835	AR129835 Sequence
12	175	17.1	2821	3	TNAF000606	AF000606 Trichoplu
13	175	17.1	2821	6	AR129836	AF129836 Sequence
14	172.5	16.9	1011	1	CTHA20770	CTHA20770 Clostridi
15	172.5	16.9	3067	1	CTCH1ACGN	Z68934 C.thermocol
16	170.5	16.7	207870	2	AC115017	AC115017 Mus muscu
17	170	16.6	5437	1	CASR69XYN2	AF036924 Caldicell
18	169	16.5	3114	8	AF093132	AF093132 Pneumocys
19	167	16.3	256635	2	AC128838	AC128838 Rattus no
20	166	16.2	1500	8	AY166602	AY166602 Magnapor
21	165.5	16.2	4291	8	AF080221	AF080221 Candida a
22	165.5	16.2	4792	8	AF254147	AF254147 Candida a
23	165.5	16.2	238046	2	AC095003	AC095003 Rattus no
24	164.5	16.1	225849	2	AC098905	AC098905 Rattus no
25	164	16.0	4675	3	DDU20608	U20608 Dictyosteli
26	163.5	16.0	110000	3	AC116957_0	AC116957 Dictyoste
27	162.5	15.9	222896	2	AC119773	AC119773 Rattus no
28	162	15.8	4284	3	DDU20661	U20661 Dictyosteli
29	161.5	15.8	252801	2	AC095211	AC095211 Rattus no
30	161.5	15.8	300029	1	AE012556	AE012556 Xylella f
31	161	15.7	9435	3	AF269242	AF269242 Plasmodiu
32	160.5	15.7	110000	2	AC109525_0	AC109525 Rattus no
33	160.5	15.7	237019	2	AC094443	AC094443 Rattus no
34	160	15.6	124636	9	AP000721	AP000721 Homo sapi
35	159.5	15.6	259772	2	AC106943	AC106943 Rattus no
36	158	15.4	9636	6	PFAPS230X	AR178096 Sequence
37	158	15.4	9636	6	AR178096	AR178096 Sequence
38	158	15.4	9636	6	PFAPS230A	L04162 Plasmodium
39	158	15.4	9654	3	AE001393	AE001393 Plasmodiu
40	158	15.4	251908	2	AC094522	AC094522 Rattus no
41	158	15.4	252301	2	AC111805	AC111805 Rattus no
42	158	15.4	4034	14	TTV1TP	X14717 Thermoprote
43	157.5	15.4	13669	14	TTV1TP	X14855 Thermoprote
44	157.5	15.4	13669	14	TTV1TP	X14855 Thermoprote
45	157	15.3	1784	8	AF254145	AF254145 Candida a

RESULT 1

ALIGNMENTS

LOCUS	CUJ9835	1201 bp	mRNA	linear	PLN 18-JAN-1996
DEFINITION	Coccidioides immitis immunoreactive cell wall protein mRNA, complete cds.				
ACCESSION	U39835				
VERSION	U39835.1	GI:1161372			
KEYWORDS					
SOURCE	Coccidioides posadasii				
ORGANISM	Coccidioides posadasii Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.				
REFERENCE	1 (bases 1 to 1201) Dugger,K.O., Villareal,K.M., Nguyen,A., Zimmermann,C.R., Law,J.H. and Galgiani,J.N. Cloning and sequence analysis of the cDNA for a protein from Coccidioides immitis with immunogenic potential Biochem. Biophys. Res. Commun. 218 (2), 485-489 (1996)				
TITLE	2 (bases 1 to 1201) Dugger,K.O. and Galgiani,J.N. Direct Submission Submitted (01-NOV-1995) Kirs O. Dugger, Infectious Diseases, University of Arizona, 1501 N. Campbell, Tucson, AZ 85724, USA Location/Qualifiers 1..1201				
JOURNAL					
REFERENCE					
TITLE					
AUTHORS					
JOURNAL					
FEATURES					
source					

CDS

```

/note="33 kDa apoglycoprotein"
/codon_start=1
/product="immunoreactive spherule cell wall protein"
/protein_id="AAC49140.1"
/db_xref="GI:1161374"
/translation="MQRFSHALALVAAGLSAQIPDIPCLANCFEVALGNDCTRLT
DFKCHSKPELPGLQITPCVEACPLDAIISNIVVDCSAGVPIIDIPVDTAAPE
PSTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPT
PTASTPAEPFPAAGSNVRASVGGIAALALAAVL"
BASE COUNT      249 a      369 c      275 g      308 t
ORIGIN

```

Alignment Scores:

Pred. No.:	1,576-51	Length:	1201
Score:	123.00	Matches:	194
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-10-081-935-4 (1-194) X CIU39835 (1-1201)

OY	1	MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu	20
Db	152	ATGAGATTCTCTCAGGCTCTCATGCTCTCGTCGCTGGCGGCGCTTCGACAGTCCAGCTTC	211
OY	21	ProAspIleProProCysAlaLeuAsnCysPheValGluAlaLeuGluAsnAspGlyCys	40
Db	212	CCAGGATCCACCTTGGCGCTTCACACGCTTCGTGAGGCTCTTCGGACAGATGGCTGC	271
OY	41	ThrArgLeuThrAspPheLeuCysHisCysSerLysProGluLeuProGlyGlnIleThr	60
Db	272	ACTGCGTTGACGCGACTTCMACTGCCACACTGCTCCAGGCTGAGCTCCCGAGGACAGATCACT	331
OY	61	ProCysValGluGluAlaCysProLeuAspIlaArgIleSerValSerAsnIleValAla	80
Db	332	CCTTCGCTTGAGAGAGGCTGCCCTTCGACGCCGATCTCCGCTCCAAACATGCTCCTT	391
OY	81	AspGlnCysSerLysAlaGlyValProIleAspIleProProValAspThrThrAlaAla	100
Db	392	GACCAAGTCTCCAAAGGCCGCTGCTCCCAATTGACATCCACAGATTGAACACACCGGCGCT	451

Oy	101	trogIuprosesrgtutrralagluProthrIagiuprothrglugluProthrIagi	120
Dd	452	CCGAGGCCATCCGAGACCGCTGAGCCACCCTGAGCCACCGAGAGCCACTGCCAG	511
Oy	121	ProthralagiuprothralagiuprothriagiuprothrglugluProthralava	140
Dd	512	CCTACGGGTAGCCACCGCTGAGCCGCACTCATGAGCCACCGAGAGACCCACTGCCTC	571
Oy	141	ProthrgluthrnglyglyglyvalProthrclythrnglyserPheethValThrclyarg	160
Dd	572	CCAACGGCACGTGGCGGTGGTCCCCA CTGGACCGGTTCCTTCA CCGTCACTGGACGA	631
Oy	161	ProthralaserThrproIagiuphaseproglIagiSerAsenValArgalaserVal	180
Dd	632	CCAACGTGCTCACCCACGCTGA GTTCCCAGGTGCTGGCTCCAAAGTCCGTGCCAGCTT	691
Oy	181	Glyglytllealaalaaleuleuenglyleualaalytleu	194
Dd	692	GGCGGCATTGCTGCTGCTCTCTCGGTCTCCCTGCTCACTCG 733	

RESULT 2	
CIU32518	CIU32518
LOCUS	CIU32518
DEFINITION	Coccidioides immitis antigen 2 (Ag2) mRNA, complete cds.
ACCESSION	U32518
VERSION	U32518.1 GI:1200179
KEYWORDS	
SOURCE	Coccidioides posadasii
ORGANISM	Coccidioides posadasii

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 1234)
Zhu, Y., Yang, C., Magee, D.M., and Cox, R.A.
Molecular cloning and characterization of Coccidioides immitis
antigen 2 cDNA
Infect. Immun. 64 (7), 2695-2699 (1996)
96294782
8698497
2 (bases 1 to 1234)
Cox, R.A.
Direct Submission
Submitted (27-JUL-1995) Texas Center for Infectious Disease.

782223

COMMENT	On Feb 24, 1996 this sequence version replaced gi:1184058.
FEATURES	Location/Qualifiers
source	1..1234

```
/organism="Coccidioides posadasii"  
/mol_type="mRNA"  
/strain="Silveira"
```

gene	1. .1234
	/gene="Ag2"
CDS	175. .759

```

spherule cell walls"
/codon_start=1
/product="antigen 2"
/protein_id="AAB18148.1"
/db_xref="GI:1200180"
/translation="MQFSSALIALVAGLSAQLPDIPPCALNCFVEALGNDGTRLT
DFKHSKPELPGQITPCVSEACPLDARISSNIVVQCSAGVPIDIPVDITAAPAP
PSAEIETAEETETPEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAE
PTASTPAEFPQAGSNVVASGGIAALLGLAAYL"
polya_site
1234

```

	/gene="Ag2"	
	/note="21 A nucleotides"	
BASE COUNT	252 a	277 g
ORIGIN	383 c	322 t

Alignment Scores:


```
* as soon as it is available and the accession number will
* be preserved.
1      49: contig of 49 bp in length
50      149: gap of unknown length
150      347: contig of 198 bp in length
348      447: gap of unknown length
448      4812: contig of 4365 bp in length
4813      4912: gap of unknown length
13062      13062: contig of 8150 bp in length
13163      13162: gap of unknown length
34712      34712: contig of 21550 bp in length
34713      34812: gap of unknown length
57164      57164: contig of 22352 bp in length
57264      57264: gap of unknown length
57265      97478: contig of 40214 bp in length.
FEATURES
source
1..97478
    /organism="Magnaaporthe grisea"
    /mol_type="genomic DNA"
    /db_xref="taxon:148305"
    /chromosome="7"
    /clone="11K22"
BASE COUNT      22756 a 24891 c 24967 g 24263 t      601 others
ORIGIN
Alignment Scores:
Pred. No.:      0.0409      Length:      97478
Score:          206.50      Matches:      65
Percent Similarity: 47.69%      Conservative: 28
Best Local Similarity: 33.33%      Mismatches: 81
Query Match:      20.19%      Indels:      22
DB:              2          Gaps:      10
US-10-081-935-4 (1-194) x AC105929 (1-97478)
QY      1 MetGlnPheSerHisAlaLeuIleAlaLeuAlaGlyLeuAlaSerAlaGlnLeu 20
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      82807 ATGAGATTCTCC---ATCATCTCTGCGCGCGCGCCGCGCTCGCTCCGAGACCTC 82751
QY      21 ProAspIleProProCysAlaLeuAlaLeuAlaLeuAlaGlyLeuAlaSerAlaGly 40
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      82750 TCTTCCCTCCCGCGCTGCGCTGCGACTGCTGAGCAGCGCTATTCTCTCCGCGCTCG 82691
QY      41 ThrArgLeuThrAspPheLysCysHisCys-----SerLysProGluLeuProGlyGln 58
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      82690 AACGCG---ACCGAGCTCGCGCTCCAGCTGCGTGAAGAAGAGACATTGCCAACGTC 82634
QY      59 IleThrProCysValAlaGluAlaCysProLeuAspAlaArgLLeSerValSerAsnIle 78
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      82633 GCTACTCTTGGCTGCTCGCCCAAGTGCACGACCTCGGCGACCTCTCCAAAGCTCGCAC 82574
QY      79 -ValValAspGlnCysSerLysAlaGlyAlaProIleAspIleProProValAspThrTh 98
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      82573 AACGGGAGAGCCCTGTGCAAGAACGTCGCCACCGCGCTGTTCTTCACTCTCT-----CC 82523
QY      98 rAlaAlaProGluProSerGluThrAlaGluProThrAlaGluProThrGluGluProTh 118
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      82522 ACGCGCGCGCCCTACCAACACCAACGAG---CGGGAATCTTCCGCGCTCGCGCCACGACAC 82466
QY      118 rAlaGluProThrAlaGluProThrAlaGluProThrHisGluProThrGluGluProTh 138
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      82465 ATGCCCTCCATCGATCGATCGCTGCGCCCAAGACCAACGAGC-ACCTGCTGCTGCTCCAC 82407
QY      138 rAlaValAlaProThrGlyThrGlyGlyVal-----ProThrGlyThrGlySerThrTh 156
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      82406 CGCTTCCCGCCACAGACACGAGCGCGGCGAGCAAGAACCCAGAGGCGACACGCGGC----- 82352
QY      156 rValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyAlaGlySerAsnVal 176
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      82351 -----GGCTCTCCCTGCGACACGCGCTGCG-----GGCTCTGCGGCGCTCGCC 82308
QY      176 lArgAlaSerValAlaGlyIleAlaAlaAlaLeuLeuGlyLeu 190
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db      82307 C---GCCACTATCGGC-----GCTGCTGCTCTTCTTGCTT 82274
RESULT 6
LOCUS      AY057052      2960 bp      mRNA      linear      INV 29-Oct-2001
DEFINITION Mamestra configurata intestinal mucin (MUCIN) mRNA, complete cds.
ACCESSION      AY057052
VERSION
KEYWORDS
SOURCE
ORGANISM
Mamestra configurata (bertha armyworm)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Noctuoidea; Noctuidae; Hadeninae; Mamestra.
1 (bases 1 to 2960)
REFERENCE
AUTHORS      Shi,X., Karkuc,T., Chamankhah,M., Alting-Mees,M., Hemmingen,S.M.
and Hegedus,D.D.
5' RACEing across a bridging oligonucleotide (BO-5' RACE)
Unpublished
2 (bases 1 to 2960)
Chamankhah,M., Karkuc,T., Shi,X., O'grady,M., Visal,S.,
Hemmingen,S.M., Alting-Mees,M., Braun,L. and Hegedus,D.D.
Direct Submision
Submitted (24-SEP-2001) Molecular Genetics, Agriculture and
Agri-Food Canada, 107 Science Place, Saskatoon, SK S7N 0X2, Canada
FEATURES
source
1..2960
    /organism="Mamestra configurata"
    /mol_type="mRNA"
    /db_xref="taxon:174822"
    /cfeature_type="midgut"
1..2960
    /gene="MUCIN"
60..2495
    /gene="MUCIN"
    /note="peritrophic membrane structural protein"
    /codon_start=1
    /product="intestinal mucin"
    /protein_id="AA117912.1"
    /db_xref="GI:16508143"
    /translation="MKTLLVLTALVQARNPEDADLTNGRLYEVAHDDCPRAVHFL
LPEVDCTKFYCEVGLAKYIEPNRCASGTFNAIEOVCHPSSGSLPGFTLAPRE
TAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTA
APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAP
BEYCHLPYICDGLNRCQPPQLPDPATPEATVSWEDTCVNDGTYTPPTVAFEG
TTSAPGNDNIGVLDNGCPVDSIIHLPEHECEKYYQCDAGKIEIRNCAKGVAFNA
AQACDMPFNPVPHCAGSAGATAAPTEADESEBZPLPYDPDSWESLPGCGVDSISHLV
PHESDCDKYVCDNGLVOLGCPAGTHFSPSOOFCTWPHBAGCEHHTGGCTTTPNGG
GSGGSTAAPVDPPTTPVAVVTSAPISDPSTASAPVPESTPVAVVTTSAPISVST
SAHPDPTPVNUSSEBZPLPNPDEDLIPNGCADPFEVDLLPHETDCDKFYCVHGEI
VEPPCAPGTHESPAUDACTWPOBAGEHSEBSSTVAPELITVAIVSTLSVAPPTAAV
PPTPTVAPVAPVDPPTTGTANPACPECLPGVNPADCKECCNVAPMAHAECDKYT
CVGDEFRVNACAGEGHFNPSLTLCPIFNAGCVRNIPQITRHVEGLMFIPIHDFNRD
VIDLEHLEMAEL"
BASE COUNT      791 a      880 c      583 g      705 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      0.0255      Length:      2960
Score:          186.00      Matches:      51
Percent Similarity: 39.75%      Conservative: 13
Best Local Similarity: 31.68%      Mismatches: 77
Query Match:      18.18%      Indels:      20
DB:              3          Gaps:      5
US-10-081-935-4 (1-194) x AY057052 (1-2960)
QY      22 AspIleProProCysAlaLeuAlaLeuAlaGlyLeuAlaSerAlaGlyCysThr 41
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      162 GACTGTCTCCCTCGTGAAGTTCATCTTCTTGCGCC-----CAGCAATACGACTGCACC 215
QY      42 ArgLeuThrAspPheLysCysHis-----Cys 50
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```



```
Db      216 AAGTTT-----TACTACTGTGAATATGCTGTAAGTATATCGAACCCAGAACTGTCT 269
Oy      51 SerLyProGluLeuProGlyGlnIleThrProCyVal-----GluGuaIaCyS 67
Db      270 TCTGCTACTGACTTCACACCGCTGAGATTCACAGTTTGTTTATCCAAAGCTTTCAGATGC 329
Oy      68 ProLeuAsp-----AlaArgIleSerValSerAsnIleValValAspGlnCySerys 85
Db      330 TCTTACCGGATTCCTCCACATTAGCACGACGTAACCTCTCTCCACCACCGCAGCTCT 389
Oy      86 AlaGlyValProIleAspIleProProValAspThrThrAlaIaProGluProSerGlu 105
Db      390 ACAGCCGACCCACTGACCTCTTACAGCCGCTCCACTGACGCTTACCGCTGACCC 449
Oy      106 ThrAlaGluProThrAlaGluProThrGluGluProThrAlaGluProThrAlaGluPro 125
Db      450 ACTGCGGCCCCCTACCGCTGACCCACGCGGCCCCCTACCGCTGACCCACCGCCCT 509
Oy      126 ThrAlaGluProThrHisGluProThrGluGluProThrAlaValProThrGlyThrGly 145
Db      510 ACTGCTGACACCCACTGACGCCCCCTACCGCAGCTCTTCTGACGCCCCACCGCTGACCC 569
Oy      146 GlyGlyValProThrGlyThrGlySerPheThrValThrGlyArgProThrAlaSerThr 165
Db      570 ACAGCAGCTCTTACTGACGCTCCACAGCGCTGCTCCACCGCAGCTCTTACTGACGCCCA 629
Oy      166 Pro 166
Db      630 CCA 632

RESULT 7
ARI40164 LOCUS ARI40164 1470 bp DNA linear PAT 16-JUN-2001
DEFINITION Sequence 1 from patent US 6207436.
ACCESSION ARI40164
VERSION ARI40164.1 GI:14482660
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1470)
AUTHORS Bj.o. Blashed, rnvad, M. BekeJund., Hatakeyama, M., Schulein, M. and
Nielsen, J. Bech.
TITL E Endo-B-1,4-glucanases from saccharothrix
JOURNAL Patent: US 6207436-A 1 27-MAR-2001;
FEATURES
source 1. 1470
location/Qualifiers
BASE COUNT 242 a 563 c 489 g 176 t
ORIGIN

Alignment Scores:
Pred. No.: 0.0143 Length: 1470
Score: 185.50 Matches: 44
Percent Similarity: 52.17% Conservative: 4
Best Local Similarity: 47.83% Mismatches: 42
Query Match: 18.13% Indels: 2
DB: Gaps: 1

US-10-081-935-4 (1-194) x ARI40164 (1-1470)

Oy      83 CysSerLySAla---GlyValProIleAspIleProProValAspThrThrAlaIaPro 101
Db      424 TGCACAGGTGCGGTGGCGGCCGACGACGAGCGAGCCCGCCGAGCCCGGAGCC 483
Oy      102 GluProSerGluThrAlaGluProThrAlaGluProThrGluGluProThrAlaGluPro 121
Db      484 ACGCCCGACCGGACGCGGAGCCGACCGGAGCCGAGCGGAGCCGAGCCCGGAGCCG 543
Oy      122 ThrAlaGluProThrAlaGluProThrHisGluProThrGluGluProThrAlaValPro 141
Db      544 ACGCCGAGCCGACCGCCGAGCCGAGCCGAGCCGAGCCGAGCCGAGCCGAGCCG 603
```

```
Oy      142 ThrGlyThrGlyGlyValProThrThrGlyThrGlySerPhe-ThrValThrGlyArgPro 161
Db      604 ACCCCGAGCCGACGATGCCGCGGTCCAGGCCGCGTACGTTCACGTCGACACCGAAC 663
Oy      161 oThrAlaSerThrProAlaGluProGlyAla 172
Db      664 CAGTGTACCGCGCTTGGCAGCGCGCCAGCGCT 697

RESULT 8
AF319782/c LOCUS AF319782 149696 bp DNA circular VRL 07-NOV-2002
DEFINITION Callicitric heline herpesvirus 3 strain Cj0149, complete genome.
ACCESSION AF319782
VERSION AF319782.2 GI:24762296
KEYWORDS
SOURCE Callicitric heline herpesvirus 3
ORGANISM Callicitric heline herpesvirus 3
REFERENCE 1 (bases 1 to 105145)
AUTHORS Cho, Y., Ramer, J., Rivallier, P., Quink, C., Garber, R. L., Beier, D. R. and Wang, F.
TITL E An Epstein-Barr-related herpesvirus from marmoset lymphomas
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (3), 1224-1229 (2001)
PUBMED 21107697
JOURNAL MEDLINE 11158621
REFERENCE 2 (bases 1 to 149696)
AUTHORS Rivallier, P., Cho, Y. G. and Wang, F.
TITL E Complete genomic sequence of an Epstein-Barr virus-related herpesvirus naturally infecting a new world primate: a defining point in the evolution of oncogenic lymphocryptoviruses
JOURNAL J. Virol. 76 (23), 12055-12068 (2002)
PUBMED 12414947
MEDLINE 12414947
REFERENCE 3 (bases 1 to 105145)
AUTHORS Wang, F.
TITL E Direct Submission
JOURNAL Submitted (12-APR-2001) Medicine, Brigham & Women's Hospital, Harvard Medical School, Channing Laboratory, 181 Longwood Avenue, Boston, MA 02115, USA
4 (bases 1 to 149696)
AUTHORS Rivallier, P., Cho, Y. and Wang, F.
TITL E Direct Submission
JOURNAL Submitted (04-JUN-2002) Medicine, Brigham & Women's Hospital, Harvard Medical School, Channing Laboratory, 181 Longwood Avenue, Boston, MA 02115, USA
REMARK Sequence update by submitter
COMMENT On Nov 7, 2002 this sequence version replaced gi.13676640.
FEATURES
source 1. 149696
location/Qualifiers
1. 149696
/organism="Callicitric heline herpesvirus 3"
/mol_type="genomic DNA"
/strain="Cj0149"
/species="host="Common marmoset"
/db_xref="taxon:106331"
/note="isolated from lymphoma"
/complement(join(138637..138818,138904..139005,139106..139219,139294..139458,139572..139661,139744..139992,140075..140167,140247..140485,959..9661)/product="C7"
/complement(join(138637..138818,138904..139005,139106..139219,139294..139458,139572..139661,139744..139992,140075..140167,140247..140485,959..9661)/product="C7"
/codon_start=1
/product="C7"
/protein_id="AA064290.1"
/db_xref="GI:24762310"
/translation="MAGHWYEVYIPGFLCPLILPSLFWICSLTLFLVGHANIVASVLFVLAMCLILAMNNTTRBPVGRSRSSMSLSVAASATATPMFASFTLSFDDGLLLFGTALVQITIVLVLMVEITVIMKRYVHFVITTLFLSPITLISVACILIOSALLIEAVVTTITVLAIPLMLPQGAEDAGTLLTLNLTALCVLILTVAIPTDAQILTVFCLPQWTLFICLIRIMICNRGKLTIRIICLKFLYGLISLSFGWVAFLKEVTLPTT
```

gene
mRNA
/gene="C1"
/join(484..721,811..897,1087..1829)
/gene="C1"
/product="C1"
/note="positional and functional similarity to EBV LMP1;
transforming gene"

CDS
/codon_start=1
/protein_id="AAK38207.1"
/db_xref="GI:13676641"

CDS
/note="similar to EBV BALF1; similar to bcl-2"
/codon_start=1
/product="ORF1"
/protein_id="AAK38208.1"
/db_xref="GI:13676642"

CDS
/note="similar to EBV BALF2; major DNA binding protein"
/codon_start=1
/product="ORF2"
/protein_id="AAK38209.1"
/db_xref="GI:13676643"

CDS
/note="similar to EBV BALF3; transport protein"
/codon_start=1
/product="ORF3"
/protein_id="AAK38210.1"
/db_xref="GI:13676644"

CDS

CDS

CDS

CDS

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:
1.25	184.00	46.81%	35.46%	14
14	19696	46.81%	35.46%	3

TLARVAGDITVEEMAAVNHIFARARFARVAREDEKRDGTAEIIFENSKYIKTLFNOT
LSSEHNALMTVYSLITGLDAGCPPEPVNAVALSYSCAAGILIPHOXFMASLW
PGIOAEWAEIETSFNSFYOPPDCTLATTOOLABYIEATITVSLYNTKOSCHLRV
DEAGLEGANPPDDIYITTESBNLILITKGMIFIDOLIALTLHLOMVDGRV"

9973..12585
/note="similar to EBV BALF4; glycoprotein B"
/codon_start=1
/product="ORF4"
/protein_id="AAK38211.1"
/db_xref="GI:13676645"

CDS
/note="similar to EBV BALF5; DNA polymerase"
/codon_start=1
/product="ORF5"
/protein_id="AAK38212.1"
/db_xref="GI:13676646"

CDS
/note="similar to EBV BALF1; glycoprotein gp64"
/codon_start=1
/product="ORF6"
/protein_id="AAK38213.1"
/db_xref="GI:13676647"

Alignment Scores:
Pred. No. : 1.25 Length: 149696
Score: 184.00 Matches: 50
Percent Similarity: 46.81% Conservative: 16
Best Local Similarity: 35.46% Mismatches: 71
Query Match: 17.99% Indels: 4
DB: 14 Gaps: 3

JOURNAL J. Biol. Chem. 272 (26), 16663-16669 (1997)
 MEDLINE 97341213
 PUBMED 9195982
 REFERENCE 2 (bases 1 to 2455)
 AUTHORS Wang, P. and Granados, R.R.
 TITLE Direct Submission
 JOURNAL Submitted (21-APR-1997) Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA

FEATURES
 source
 1..2455
 Location/Qualifiers
 /organism="Trichoplusia ni"
 /mol_type="mRNA"
 /db_xref="taxon:7111"
 38..2404
 /product="insect intestinal mucin IIM14"
 /protein_id="AAC47556.1"
 /translation="MIKTLFLALGLVAPREVSDAEKNPALHEHPDPPAEQHWL
 LPHEDCTKFEYCEYGLKFIAPDCAEPFKSAOTCVAAALAGCTLPAPATTOA
 PATTOAPTTTOAPTTTOAPTTTOAPTTTOAPTTTOAPTTTOAPTTTOA
 PTTTOAPTTTOAPTTTOAPTTTOAPTTTOAPTTTOAPTTTOAPTTTOA
 AATTGATTTGVPAPTSAPVWPPICELLPNCSPADFDIHLIPDKICNLFYCCSNY
 TEORCEGLYFNPYVGRCDSPANVECDGEISAPVTEGNEDEDIDIDGLDNGCPA
 NEFIDWILPHGNRCDKYQCCHGNLYERRCGAGHFSFELQCCDIELVCGTLPGES
 EYEDVEDACTGWYCEPTPIEMEPILPNCADPSIDILLPHESDCGOYLQCVGORTA
 RPCGNLHFSFATOSCEPVTAGCOVECSNDCSTAAPTTAAPTAAPTAAPT
 AASTVVPAPATPATAVPVPTTAIPAPAPATAPTAAPTAAPTAAPTAAPT
 AAPTVAPEIPITVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT
 AAPTVAPEIPITVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT
 AAPTVAPEIPITVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT
 APTTDDDDIDPDLNDPNDIPNCEBCNVLPMHAPCDKWCNDQNOVLVVCSEGLQFNP
 TTKICDFACVNCVRSNIOMSEYEGVDFIPNNKIDEDIRQALNFEL"

CDS

BASE COUNT 592 a 961 c 479 g 423 t

ORIGIN

Alignment Scores:
 Pred. No.: 0.0939 Length: 2455
 Score: 175.00 Matches: 56
 Percent Similarity: 37.36% Conservative: 12
 Best Local Similarity: 30.77% Mismatches: 72
 Query Match: 17.11% Indels: 42
 DB: 3 Gaps: 7

US-10-081-935-4 (1-194) x TNAF00605 (1-2455)

QY 25 ProCyalaLeuabnCySpheValGluAlaLeuGlyAsnAapGlyCysThrArgLeuThr 44
 DB 1418 CCGGCCACACAGTCCTGT-----GAGTCTCTGTGACCCCTGGTGC-----CAA 1462
 QY 45 AspPheLysCys-----HisCysSerLysProGluLeuProGlyGlnIleThr 60
 DB 1463 GTTTTCGAGTGCATTCGACAAACAGTGCACATCGACTGCTGCCCGGACAGCTGCCTCA 1522
 QY 61 ProCyalaGluGluAlaCysPProLeuAapAlaArgIleSerValSerAenIleValI 80
 DB 1523 ACCGCTGCCCAACGGCTGCCCAACGGCTGCCCAACCTCCGACCCCTCACCCTGGTGC 1582
 QY 81 AspGlnCysSerLysValIaGlyValProIleAapIleProProValAapThrThrAla 100
 DB 1583 CCACCTGCAAGCCACCCGCACTGCAAGCCCACTGCACT-----ACAACCGCAATT 1636
 QY 101 ProGluProSerGluThrAlaGluProThrAlaGluProThrGlu----- 115
 DB 1637 CCTACTCCGAGCCCAACGGCTGCCCAACGGCTGCTACTACTGCTGCCCTGAAATCC 1696
 QY 116 -----GluProThrAlaGluProThrAla----- 123
 DB 1697 CCAACCACTGTACAGTACACCTACTGTCTCCACCGCAGCCCTACTACTGCTGTGC 1756
 QY 124 -----GluProThrAlaGluProThrAlaGluProThrAlaGluProThr 134
 DB 1757 CCGAATCCCAATCACTGTCATCATGAGCCCTTACCGCTGCCCAACGGCTGCCCTCACC 1816

QY 135 GluGluProThrAlaValProThrGlyThrGlyGlyValProThrGlyThrGlySer 154
 DB 1817 GCTGCCCAACCGCAGCCCTTACTACTGCTGTCTCCAAATCCCACT----- 1864
 QY 155 PheThrValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyAlaGlySer 174
 DB 1865 ---ACTGTACATCACCACCTACTGCTGCCCACTACCGAGCAGCCTGCCCAACACC 1921
 QY 175 AsnVal 176
 DB 1922 ACAGTC 1927

RESULT 11

ARI129835
 LOCUS ARI129835 2455 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 1 from patent US 6187558.
 ACCESSION ARI129835
 VERSION ARI129835.1 GI:14117732
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1. (bases 1 to 2455)
 AUTHORS Granados, R.R. and Wang, P.
 TITLE Invertebrate intestinal mucin cDNA and related products and methods
 JOURNAL Patent: US 6187558-A 1 13-FEB-2001;
 FEATURES
 source
 1..2455
 Location/Qualifiers
 /organism="unknown"

BASE COUNT 592 a 961 c 479 g 423 t

ORIGIN

Alignment Scores:
 Pred. No.: 0.0939 Length: 2455
 Score: 175.00 Matches: 56
 Percent Similarity: 37.36% Conservative: 12
 Best Local Similarity: 30.77% Mismatches: 72
 Query Match: 17.11% Indels: 42
 DB: 6 Gaps: 7

US-10-081-935-4 (1-194) x ARI129835 (1-2455)

QY 25 ProCyalaLeuabnCySpheValGluAlaLeuGlyAsnAapGlyCysThrArgLeuThr 44
 DB 1418 CCGGCCACACAGTCCTGT-----GAGTCTCTGTGACCCCTGGTGC-----CAA 1462
 QY 45 AspPheLysCys-----HisCysSerLysProGluLeuProGlyGlnIleThr 60
 DB 1463 GTTTTCGAGTGCATTCGACAAACAGTGCACATCGACTGCTGCCCGGACAGCTGCCTCA 1522
 QY 61 ProCyalaGluGluAlaCysPProLeuAapAlaArgIleSerValSerAenIleValI 80
 DB 1523 ACCGCTGCCCAACGGCTGCCCAACGGCTGCCCAACCTCCGACCCCTCACCCTGGTGC 1582
 QY 81 AspGlnCysSerLysValIaGlyValProIleAapIleProProValAapThrThrAla 100
 DB 1583 CCACCTGCAAGCCACCCGCACTGCAAGCCCACTGCACT-----ACAACCGCAATT 1636
 QY 101 ProGluProSerGluThrAlaGluProThrAlaGluProThrGlu----- 115
 DB 1637 CCTACTCCGAGCCCAACGGCTGCCCAACGGCTGCTACTACTGCTGCCCTGAAATCC 1696
 QY 116 -----GluProThrAlaGluProThrAla----- 123
 DB 1697 CCAACCACTGTACAGTACACCTACTGTCTCCACCGCAGCCCTACTACTGCTGTGC 1756
 QY 124 -----GluProThrAlaGluProThrAlaGluProThrAlaGluProThr 134
 DB 1757 CCGAATCCCAATCACTGTCATCATGAGCCCTTACCGCTGCCCAACGGCTGCCCTCACC 1816
 QY 135 GluGluProThrAlaValProThrGlyThrGlyGlyValProThrGlyThrGlySer 154
 DB 1817 GCTGCCCAACCGCAGCCCTTACTACTGCTGCCCACTACCGAGCAGCCTGCCCAACACC 1864

```

RESULT 12
TNAF000606      2821 bp      mRNA      linear      INV 28-JUN-1997
LOCUS           Trichoplusia ni insect intestinal mucin IIM22 mRNA, complete cds.
DEFINITION      AF000606
ACCESSION       AF000606.1  GI:2224920
VERSION
KEYWORDS
SOURCE
ORGANISM
Trichoplusia ni (cabbage looper)
Trichoplusia ni
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
Noctuoidea; Noctuidae; Plusiinae; Trichoplusia.
REFERENCE
1 Wang, P. and Granados, R.R.
Molecular cloning and sequencing of a novel invertebrate intestinal
mucin cDNA
J Biol. Chem. 272 (26), 16663-16669 (1997)
J9741213
MEDLINE
PUBMED
9195982
REFERENCE
2 (bases 1 to 2821)
Wang, P. and Granados, R.R.
Direct Submission
Submitted (21-APR-1997) Boyce Thompson Institute, Cornell
University, Ithaca, NY 14853, USA
FEATURES
source
1..2821
/organism="Trichoplusia ni"
/mol_type="mRNA"
/db_xref="taxon:7111"
26..2449
/codon_start=1
/product="insect intestinal mucin IIM22"
/protein_id="AAC47557.1"
/db_xref="GI:2224921"
translation="MIKTLPLFLALGLVAARREVSDAEKNPALHEPHDPDCEPAECOHML
LPHRYDCKRKYCEVCGKEFLAPDCACTERKESFQNCVNAALGCTLPGPAPPTQQA
PATTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQAP
PTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQAPPTTQ
AATPAAATPGVPAPTSAIPWMPPIICELPNCGPCADFDHLIIPDKCNLFYQCSNQR
TFEORCEPGLFVNPVORCDSPANVECDGEISPAHPVTEGDEDDIDIGLDNCGPAA
NEFLIDMLPHNCRGCDKRYOCVGHNLVERSCAGHFSFLQCOCHILEVGCTLLPGSES
EUEVDEDACGACGWCPTREPIEMEPPLNCGCPADPSIDHLIPHESDCGOYLQCVHQTTLA
RPGCGNLHFSPATOSCESPYTAGCOVFECDSDDNCTSAATAPTAAPTAAPTAAPTAAPT
AAPTVAPEIPIVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT
AAPTVAPEIPIVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT
AAPTVAPEIPIVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT
VAPAPNTAAVTTTASAPATPEDDIDPLPNPINDPINCVECVNLPMAHADCDKMYW
CDGNQNVLYVSEGGQFPPTTKTCDFAFNCCVANSNIQMSSEYEGVQVFIPMNLDEEL
IRQALNFEFL"
BASE COUNT      710 a      1030 c      529 g      552 t
ORIGIN
Alignment Scores:
Pred. No.:      0.107      Length:      2821
Score:          175.00      Matches:      56
Percent Similarity: 37.36%      Conservative: 12
Best Local Similarity: 30.77%      Mismatches: 72
Query Match:    17.11%      Indels:      42
DB:             3          Gaps:      7
US-10-081-935-4 (1-194) x TNAF000606 (1-2821)
Oy 25 ProcysALaLeuAsnCysPheValG1uAlaLeuG1yAsnAspG1yCysThrArgLeuThr 44

```

Db	1406	CCTGCACACAGCTCTCTGT-----GAGTCTCTCTGTACACCCCTGTTGC-----CAA	1450
Qy	45	AspPheLysCys-----HisCysSerLysProGluLeuProGlyGlnIleThr	60
Db	1451	CTTTTGAGAGTGCATTCTGTACACACAGTGTACATTCAGTCTGTGCGCCGACAGCTGTCCA	1510
Qy	61	ProCysValGluGluAlaCysPheLeuAspAlaArgIleSerValSerAsnIleValIle	80
Db	1511	ACCGCTGCCCCCAACGGCTGCTGCCCAACGGCTGCCCCCAACCTGCCACCTCTGTC	1570
Qy	81	AspGlnCysSerLysValIleGlyValProIleAspIleProProValAspThrThrAlaIle	100
Db	1571	CCACTGTCAACGGCACCAGCACTGACAGCCCACTGCTTCCACT-----ACAAACCGCAATT	1624
Qy	101	ProGluProSerGluThrAlaGluProThrAlaGluProThrGlu-----	115
Db	1625	CTTACTCGGGCCCCCAACCGCTGGCCCCCAAGCTCTTACTGTGTCGCCCTGAATCC	1684
Qy	116	-----GluProThrAlaGluProThrAla-----	123
Db	1685	CCAACCACTGTCAAGTACAGTACCCTACTGCTGTCCACCGACGCCCTACTACTGTGTC	1744
Qy	124	-----GluProThrAlaGluProThrAlaGluProThrIleGluProThr	134
Db	1745	CTGAATATCCCAATCACTGTCACATCAGCGCTTACATCGCTGCCCCACCGCTGCCCCAC	1804
Qy	135	GluGluProThrAlaValProThrGlyThrGlyGlyValProThrGlyThrGlySer	154
Db	1805	GCTGCCCCCAACCGACGCCCTTACTACTGCTGTCTCCAGAAATCCCACT-----	1852
Qy	155	PheThrValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyAlaGlySer	174
Db	1853	---ACTGTCAATACACCACTACTGCTGTGCCCCCACTACCGACGACGACTGCCCAACACC	1909
Qy	175	AsnVal 176	
Db	1910	ACAGTCT 1915	
RESULT 13			
LOCUS	ARI29836	2821 bp	DNA linear PAT 16-MAY-2001
DEFINITION	Sequence 2 from patent US 6187558.		
ACCESSION	ARI29836		
VERSION	ARI29836.1	GI:14117733	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 2821)		
TITLE	Granados, R.R. and Wang, P.		
JOURNAL	Invertebrate intestinal mucin cDNA and related products and methods		
FEATURES	Patent: US 6187558-A, 2, 13-FEB-2001;		
	Location/Qualifiers		
	1..2821		
BASE COUNT	710 a	1030 c	529 g 552 t
ORIGIN	/organism="unknown"		
Alignment Scores:			
Pred. No.:	0.107	Length:	2821
Score:	175.00	Matches:	56
Percent Similarity:	37.36%	Conservative:	12
Best Local Similarity:	30.77%	Mismatches:	72
Query Match:	17.11%	Indels:	42
DB:	6	Gaps:	7
US-10-081-935-4 (1-194) x ARI29836 (1-2821)			
Qy	25	ProCysAlaLeuAsnCysPheValGluAlaLeuGlyAsnAspGlyCysThrArgLeuThr	44
Db	1406	CTGTCCACACAGCTCTGT-----GAGTCTCTCTGTACACCCCTGTTGC-----CAA	1450
Qy	45	AspPheLysCys-----HisCysSerLysProGluLeuProGlyGlnIleThr	60

TITLE Direct Submission
JOURNAL Submitted (27-JAN-1996) Fuchs K.P., Institute for Microbiology,
Technical University of Muenchen, Arcistr. 16, Muenchen, Bavaria,
Germany, D-80290

REMARK 3 (bases 1 to 3067)

AUTHORS Fuchs, K.P.

TITLE Direct Submission

JOURNAL Submitted (13-DEC-2001) Fuchs K.P., Institute for Microbiology,
Technical University of Muenchen, Arcistr. 16, Muenchen, Bavaria,
Germany, D-80290

COMMENT On Dec 23, 2001 this sequence version replaced gi:1418679.

FEATURES
Location/Qualifiers

source
1..3067
/organism="Clostridium thermocellum"
/mol_type="genomic DNA"
/strain="DSM 1237"
/db_xref="taxon:1515"
/clone_lib="pKL133"
/clone_1ib="lambda 1059"
137..140
/note="putative promoter region"
159..164
/note="putative promoter region"
174..1633
/gene="ChiA"
174..178
/gene="ChiA"
185..1633
/gene="ChiA"
/EC_number="3.2.1.14"
/function="hydrolysis of beta-1,4 linkages of chitin"
/note="estimated molecular mass 55 kDa"
/codon_start=1
/transl_table=1
/product="Chitinase"
/protein_id="CAA93150.1"
/db_xref="GI:1418680"
/db_xref="SPRMBL:Q59326"
/translation="MKKIPLMILSLAISIIPLSLHPTLSYADDSLPTRKIYGFAMENI
YLENNYVESDIPMDWYTHINAYFAKIEIGIALIDKMAIQQFPAILOTPIRGHG
QLIKKEQYHPHVKTLISVGWTESKYFSDVALTESNNTFADSCVEIRIYAPFGVDI
DMEYVSGMPENIRPEDKONFTLLKCLREKDAAGAEKGKYLITIAAPAGSPNI
KNEPEIYHOYLDIFINIMTYDYGSGMENVAHLAPLYMNPDPSPYPERKKNFNDWV
KEYRIKGVPAEKINIGVYVAAQWQVNGGICGLFGTSSKPLSTGTHYISNLKSPD
LGFTRYDEYAMVPYLYNRPESATYISYEDELSTKCKDYIENNLGSIIMTWELSGITA
EGGTTLSVLYDSFTSFKDELYGDLNDGKVNSDLAILKRYMLRAISDFPIPGRL
ADLNRDGNVNSTDYSILKRYILKAIIDNIPVDV"
185..265
/gene="ChiA"
/note="putative"
1424..1633
/note="2x29 AA repeat"
1795..1833
/note="palindrome"
1849..2772
/gene="ORF2"
1849..2772
/gene="ORF2"
/note="similarity to typeIII-cellulose-binding proteins"
/codon_start=1
/transl_table=1
/product="putative substrate binding protein"
/protein_id="CAA93151.2"
/db_xref="GI:17977823"
/db_xref="SPRMBL:Q46392"
/translation="MKWYKSKVRQVILKITLPVVLVLTSLVYVYFLTNRRGSRD
ADGTPDPSVSPSIAINTPSGDTASAGLVDPNTFLTEHTNAPTPDDITPTPT
LEPTPEATSTPTPTPTSPKTPKPTAPATPTPTPTPTPTPKPKOKPEKGP
ITVOYKGDSSVTAIYPIKFTINNGTSVLSGLIIRYYTTEGNEENFWNEFT
RDSQVYCTPVKSKPKKNADHYLEIGYDPAKSLKRESVELKVNKNGMTTTSLSM
TTSYRNVRNRFITNMIILTVYLSGKLVGKEP"
2831..2861
terminator

BASE COUNT 998 a 559 c 658 g 852 t
ORIGIN
Alignment Scores:
Pred. No.: 0.161 Length: 3067
Score: 172.50 Matches: 39
Percent Similarity: 55.17% Conservative: 9
Best Local Similarity: 44.83% Mismatches: 38
Query Match: 16.86% Indels: 1
DB: 1 Gaps: 1
US-10-081-935-4 (1-194) x CTCH1AGEN (1-3067)

QY 81 ASpglncyseryserysagly---ValprolleasplreproprovalasprthrAla 99
Dl 2050 GATGCAACTGACATCGGACGACTTGATCGACATCATTTTTCAGCGAGCACCAAT 2109
QY 100 AlaprogIuprosergIutrralagIuproThralagIuproThrgIugIuproThrala 119
Db 2110 GCGCCGACTCAACAGATGACATTACCGCTTACACCTTACTCTTGAACCAAGCCT 2169
QY 120 GluproThralagIuproThralagIuproThHieGluproThrgIugIuproThAla 139
Db 2170 GAACCAACTGCTACACGACATCATCACCAACATTACCGCGGACGTCMAAACCAAGCC 2229
QY 140 ValproThrgIythrGlygIyValproThrgIyThrgIyserpThrrvalThrgIy 159
Db 2230 AAGCCGACTCGACACGACGCAAAACGACCGCGGACCTACGACACCAACCAAGC 2289
QY 160 ArgProThralaserThrPro 166
Db 2290 AAACCTGCACAAAAGACACCT 2310

RESULT 16
AC115017/c
LOCUS
DEFINITION
AC115017 207870 bp DNA linear HTG 23-APR-2003
Mus musculus clone RP24-220A11, WORKING DRAFT SEQUENCE, 12 ordered
pieces.
AC115017 GI:30017911
VERSION
HTG; HTGS PHASE2; HTGS DRAFT.
KEYWORDS
Mus musculus (house mouse)
SOURCE
ORGANISM
Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
Unpublished
2 (bases 1 to 207870)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,S., Barta,N., Baetien,V., Bloom,T., Bognulavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campolano,A., Chang,J.,
Chazaro,B., Choquel,Y., Collins,M., Collins,M., Collymore,A.,
Cook,A., Cooke,P., DeArrelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margulis,N., Mathews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menus,L.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunhahng,P., Pierre,N., Pollara,V.,
Raymond,C., Retter,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rossetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Teafay,S., Theodore,J.,
Tophan,K., Travers,M., Travis,N., Tigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE	Direct Submission
JOURNAL	Submitted (14-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE	3 (bases 1 to 207870)
AUTHORS	Birren,B., Nusbaum,C., Lander,E., Aboueleil,A., Allen,N.,

TITLE	Direct Submission
JOURNAL	Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Apr 17, 2003 this sequence version replaced qi:11338561.

```

FEATURES
*      169838 169937: gap of 100 bp
*      169938 207870: contig of 37933 bp in length.
      Location/Qualifiers

```



```

IVFGSGNGCYDNCYGVSPYTTIGSIDVRGIRHYFSEOCSSVLATVSGSIV
TNARITVTDVGEAGCTVHSGSSASTPIAGVIALVLSVNPNTMHDIOGLIVESAP
FSLDYGMEKLPSCRYYHYHYEGKLDAYMEVFAARNFKLNJFARNSVMTIYNKFE
SENNGHITDKFENLKAIPDYNYFGKLEVSVLVFOHAKGSEINTSSQYTMNT
HRLHDKNGCYVMTSTVYKMEPIVGEVTTIDEDKITNLGERTYVWOLHFGESA
DLTKPISIVPTFTRSSSPIPGEPVYLPTPSATISPSVSPQPTSESTSPPTQPO
PTPOPTSEPTSEPTSEPTSEPTQPTQPTSEPTSEPTSEPTQPTQPTQPTQPTQ
PAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQAPQ
STSEPTPRPPTQPTSEPTSEPTSEPTQPTQPTQPTQPTQPTQPTQPTQPTQ
PAPKPTPPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK
ASHTSEFKRLLLISLILSLILILILILILILILILILILILILILILILIL
MSOETEKATMLRNQOVTSISDSDIEDPMYERILRS"

BASE COUNT      1055 a      623 c      515 g      921 t

ORIGIN
Alignment Scores:
Pred. No.:      0 262      Length:      3114
Score:          169.00      Matches:      32
Percent Similarity: 61.33%      Conservative: 14
Best Local Similarity: 42.67%      Mismatches: 25
Query Match:    16.52%      Indels:      4
DB:             8          Gaps:      1

US-10-081-935-4 (1-194) x AF093132 (1-3114)

QY      93  ProProvaAspThrAlaAlaProGluProSerGluThrAlaGluProThrAlaGlu 112
      Db  2008  CCACACATCTGACGACCATCTGACCAACACCTCAACCAACCAACCAACCTCAG 2067

QY      113  ProThrGluGluProThrAlaGluProThrAlaGluProThrAlaGluProThrAlaGlu 132
      Db  2068  CCACACATCTGACGACCAACATCTGACCAACATCTGACCAACATCTGACCAACCAACCAACCTCA 2127

QY      133  ProThrGluGluProThrAlaValProThrGlyThrGlyGlyGlyValProThrGlyThr 152
      Db  2128  CCACACCTCAGCCACCATCTGACCAACATCTGAC-----CCACATCTGAG 2175

QY      153  GlySerPheThrValThrGlyArgProThrAlaSerThrProAla 167
      Db  2176  CCACACATCTGACCAACACCTCAACCAACCAACCAACCTCAACCAACCA 2220

RESULT 19
AC128838/c 256635 bp DNA linear HTG 19-NOV-2002
LOCUS      Rattus norvegicus clone CH230-395E20, *** SEQUENCING IN PROGRESS
DEFINITION
** 4 unordered pieces.
ACCESSION  AC128838
VERSION    AC128838.3 GI:25073624
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE     Rattus norvegicus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus;
1 (bases 1 to 256635)
Muzny,D,Marie, Metzker,M, Lee, A, Abramson, S, Adams, C, Alder, J,
Allen, C, Allen, H, Albrooks, S, Amin, A, Anguiano, D,
Anyalelebechi, V, Ayagi, A, Ayodeji, M, Baca, E, Baden, B,
Balwin, D, Bandaranaike, D, Barber, M, Barnstead, M, Benham, F,
Biswal, K, Blair, J, Blankenburg, K, Blyth, P, Brown, M,
Bryant, N, Buhay, C, Burch, P, Butrell, K, Calderon, E,
Cardenas, V, Carter, K, Cavazos, I, Ceasar, H, Center, A,
Chacko, J, Chavez, D, Chen, G, Chen, R, Chen, Y, Chen, Z, Chu, J,
Cleveland, C, Cockrell, R, Cox, C, Coyle, M, Cree, A, D'Souza, L,
Davila, L, Davis, C, Davy-Carroll, L, De Anda, C, Dederich, D,
Delgado, O, Denson, S, Deramo, C, Ding, Y, Dinh, H, Divya, K,
Dreber, H, Dugan-Rocha, S, Dunn, A, Durbin, K, Duval, B, Eaves, K,
Egan, A, Escotto, M, Eugene, C, Evans, C, A, Falls, T, Fan, G,
Fernandez, S, Finley, M, Flagg, N, Forbes, L, Foster, M, Foster, P,
Frazer, C, M, Gabisi, A, Ganta, R, Garcia, A, Garner, T, Garza, M,
Gebregorgis, E, Geer, K, Gill, R, Grady, M, Guerra, M, Guevara, W,
Gunaratne, P, Haaland, W, Hamil, C, Hamilton, C, Hamilton, K,
Harvey, Y, Haylak, P, Hawes, A, Henderson, N, Hernandez, J,

```

```

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hughes, M.,
Hollins, B., Howelle, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensunha, L., Loulesed, H., Lozada, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Matinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Mllogavljovic, A., Miner, G., Ming, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokeneme, O., Okunolu, G., Olarinmoye, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Plankoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L.,
Puzo, M., Quirez, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, R., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smales, D.,
Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wietzyk, R., Wooten, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von
Weinstock, G., and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 256635)
Worley, K.C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 256635)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23915267.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KAVH
Center clone name: CH230-395E20
----- Summary Statistics
Assembly program: Phrap, version 0.990329
Consensus quality: 199691 bases at least Q40
Consensus quality: 203386 bases at least Q30
Consensus quality: 205850 bases at least Q20
Estimated insert size: 203171; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

```


Db		496	GCCGACGACTGCATCAGCGTCGGCTGTGGCATTCGCCGTCTGGCGCCAGCCGTCTT	555		
Oy		32	Va Glua AlaLeuGIYAsnAspGLYCysThrArgLeuThraSPheLysCyshiCyS---50 ::: :::			
Db		556	CTCAACGGGCGCCCGCACCATTTGGCTGCTCCGCG--ACCGACTTCAAAGTGCCAGTGGCCAG	612		
Oy		51	SerLysProGluLeuProGlyGlnIleThrProCysValGIuGluAlaCySpro-----68 ::: :::			
Db		613	CAGCAGGCCAAGATTGTTGGCCGCCCTCGAGACTCGTGCATAAAAGAAGCTGCCAAGTCT	672		
Oy		68	-----68			
Db		673	GAGTTCAAAAGACCATTGACGATCTGACAAGGGTGAGTTGTCTGCCGCGCAAGATGC	732		
Oy		69	LeuAspAla-ArgIleSerValSerPasnIleValValaspGInCYserLysAlaGlyVa8 ::: :::			
Db		733	GCAAAATACCTGCCATGCATACAAACAAAATGTACTTAACACAAAGAAGACTGTGCAGTGC	792		
Oy		88	IProIleasPIeProProValasPTThrTrAlaIaProGIUpProSergIUThrAlaGI108 ::: :::			
Db		793	GCTTCTGGAGGTCCTCCGCTTCAAAACAACGCGGGGGCGCCGGA-----AT837			
Oy		108	uProThralaGIUpProThrgluUpProThralaGIUpProThralaGIUpProThr---126 ::: :::			
Db		838	ACCgGCAACCCGCTCCAGCTTCATC-CAGGGCCCCAGCTGCATGCTCTCCACACACAC896			
Oy		127	---AlaGIUpProThrHISglUProThrcIuGIUpProThrAlaValProThngLYThGI145 ::: :::			
Db		897	CGTCCGACGCCCGACCGCAACACGCTGTGGCCGACCTCGCGGCGCCACT-----948			
Oy		145	yGIyGIyValProThrgLyThrgLysPheThrValThrgIaYargProThralaSerTh165			
Db		948	-----948			
Oy		165	rProIlaGIUpHeProGIYAlaGlySerAsnValArghLa-----178			
Db		949	-----GCCGCCGCAACCACTGGCCCTGTGCAGTGCAGCAGCATCTGTGT989			
Oy		179	---SerValGIyGIYIleAlaAlaAlaLeuLeuGIYLeu190 ::: :::			
Db		990	TGCTGCCGTGGGTGTGGCGCTTGGGTAGCCCTTGCGTGTG1029			
RESULT 21						
LOCUS	AF080221	4291 bp	DNA	linear JUN 09-AUG-1998		
DEFINITION	Candida albicans mycelial surface antigen precursor (CSAI) gene, complete cds.					
ACCESSION	AF080221					
VERSION	AF080221.1	GI:3406797				
KEYWORDS						
SOURCE						
ORGANISM	Candida albicans					
	Candida albicans					
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.					
REFERENCE	1 (bases 1 to 4291)					
AUTHORS	Bourbonnais,Y., Lamarre,C. and Deslauriers,N. Expression cloning of the Candida albicans CSAI gene encoding a mycelial surface antigen with similarity to the Coccidioides imitis antigen 2					
TITLE	Unpublished					
JOURNAL	2 (bases 1 to 4291)					
REFERENCES	Bourbonnais,Y. and Deslauriers,N. Direct Submission					
AUTHORS	Submitted (24-JUL-1998) Biochemistry, Universite Laval, Cite Universitaire, Ste-Foy, Que GIK 7P4, Canada					
JOURNAL	Location/Qualifiers					
TITLE	1..4291					
FEATURES	/organism="Candida albicans" /mol_type="genomic DNA" /strain="ATCC32354" /db_xref="ATCC:32354" /db_xref="taxon:5476" 515..4126					
source	gene					

[illegible]

Oy	34	AlaLeuGlYasnAspGlyCyethrArgLeuThrAspPheIscyGhStCysSerIySpro	53
Db	716	AAcAcGtGtGAGACCCCATGTC--TACTGGATrCTGGGtGtTGTATTAAGCCA	772
Oy	54	GlUeubProGlyGlnIleThrProCYVaIglUgluaIaCYsProLeuAspAlaArgIle	73
Db	773	ACATTtGCTGtGtGCATTGtTCTTGTATTGtGtGAGAAGtGTAAAGCCAAAGCtTGTt	832
Oy	74	SeValSerAsnIleValaIaSpGlnCYsSerIySaGly-----	87
Db	833	TCTCTCAAGtTGGGAACtTCCATTtGTtCCCTTGtGtGtGtGATCCACTACTG	892
Oy	88	--ValProIleAspIle-----	92
Db	893	ATGtGtGCTCGAAATGTCGACAGACGtTTAAAGtCTCTCCACTGCTGTtGCATGtCT	952
Oy	93	-----ProPovaIaSpThrThraIaAlaProGluProSerGluThrAlaGluProThr	110
Db	953	TCTGAACACAGtTtGAAACATtCTTGAAACCACTGATCTTCTCAgTGTtGAATCT	1012
Oy	111	AlaGluProThrGluGluProThrAlaGluProThrAlaGluProThrAlaGluProThr	130
Db	1013	TCTCAACtGCTGAACACCTCACTCACTGAACCTCTGACACTTCATCATCTGAACCTG	1072
Oy	131	HisGluProThrGluGluProThrAlaValProThrGlyThrGlyGlyValProThr	150
Db	1073	GAGACTTCATCGAAACATCATCGAACAACtCTCTTCATCTGAACCTGCTGAACtTCA	1132
Oy	151	GlyThrGlySerPheThrValThrGlyArgProThrAlaSerThrProAlaGlu-----	168
Db	1133	TCAGAAAGATtTTtTACATACACTTCAGCCCA-----TCAACTCTGAAGATAAACCA	1186
Oy	169	-----PheProGlyAlaGlySerAsnValaArgAlaSerValGlyGlyIleAla	184
Db	1187	TACACCATCTACCCAAAGtTGtGCCAAGACT-----GCTTCTATCAATGtTtTGTCT	1237
RESULT 22			
LOCUS	AF254147	4792 bp	DNA linear PLN 21-SEP-2000
DEFINITION			Candida albicans putative cell wall protein (WAP1) gene, complete cds.
ACCESSION	AF254147		
VERSION	AF254147.1	GI:9963991	
KEYWORDS			
SOURCE			
ORGANISM			Candida albicans
REFERENCE			Candida albicans
AUTHORS			Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
TITLE			1 (bases 1 to 4792)
REFERENCE			Braun, B.R. and Johnson, A.D.
AUTHORS			TUP1, CPH1 and EFG1 make independent contributions to filamentation in Candida albicans
TITLE			Genetics 155 (1), 57-67 (2000)
JOURNAL			
MEDLINE			
PUBMED			10790384
REFERENCE			2 (bases 1 to 4792)
AUTHORS			Braun, B.R., Head, W.S., Wang, M.X. and Johnson, A.D.
TITLE			Identification and characterization of TUP1-regulated genes in Candida albicans
JOURNAL			Genetics 156 (1), 31-44 (2000)
MEDLINE			
PUBMED			20435724
REFERENCE			10978273
AUTHORS			3 (bases 1 to 4792)
TITLE			Braun, B.R. and Johnson, A.D.
JOURNAL			Direct Submision
MEDLINE			
PUBMED			Submitted (11-Apr-2000) Microbiology, UC San Francisco, 513
REFERENCE			Parnasus Ave, S-410, San Francisco, CA 94143-0414, USA
AUTHORS			Location/Qualifiers
TITLE			1..4792
JOURNAL			/organism="Candida albicans"
MEDLINE			/mol_type="genomic DNA"
PUBMED			/db_xref="taxon:5476"
FEATURES			
source			

gene
CDS

gene	1175..4231
	/gene="WAP1"
CDS	1175..4231
	/
	"...31"

```
/gene="Wap1"
/note="contains predicted signal sequence and GPI addition
sequences, indicating localization in the cell wall;
contains four cysteine-rich repeats similar to the single
repeat of Rbt5p and to Coccidioides immitis proline-rich
antigenic protein PRA; Wap1p"
```

```

/codon_start=1
/transl_table=12

```

```
/product="putative cell wall protein"  
/protein_id="AAG09792.1"  
/db_xref="GI:9963992"
```

```

/trāns]ac̣ion="MLPSIVISIVLASFVSÆSSITEAPITTTAEDNPYTTIPSVAKTA
SINGPADRIYDOLPECAKPMFONTGVTPEPPYWDTGCLCIMPFFAGAIΓSCIAEKCKG

```

QDVVSATSLGTSICSVAGVWDPMVMVPANVOSSLAAATAVASSSEOPVETSSBPAGS
SQVSSSQPAETSSSEPAETSSSEPAETSSSEEDPASEPAETSSSEESTITAPS
TBNBDVYTESVCUKTASINCEADPIYNOI.BEACAKPMEOINTGQVPCPYMDTGCLCIM

PTFAGATGSCIAEKCKGQDIVSATSLGTSICSVAGWDPYNNVPANVQSLSLAATAV
 LKEDNFIATLSTVNHANRHNADALDGLDGLDGLDGLDGLDGLDGLDGLDGLDGLD
 NEEQDNVEEVEEEDFACSCQUCVSESCQAFSESCQAFSESCQAFSESCQAFSESCQAF
 SSITSAPDSSAT

SSSSTTSTFIRTA SINGPADKLYDLPECAKPCMFMONTGITPCPYWDAGCLVMPQFA
SSSSTTSTFIRTA SINGPADKLYDLPECAKPCMFMONTGITPCPYWDAGCLVMPQFA

GAIGSLVADCKKQD1VSV15UG1SVCSVAGVNAFTIMLPLFVKNSSLSVAVATIAVFT13G
SASSTASQEPSETSSSEQEPSETASQOPAEITSSEESTITSAPSTPEDNPYTIYPSVAKT
SASSTASQEPSETSSSEQEPSETASQOPAEITSSEESTITSAPSTPEDNPYTIYPSVAKT
SASSTASQEPSETSSSEQEPSETASQOPAEITSSEESTITSAPSTPEDNPYTIYPSVAKT

ASINGFADRIYDOLPECAKPMFQUNIGVLECPIMDIGLCLIMPIFAGALGCLAEKCK
GQDVSAATSLGSSICSVAGVMDPYMMLPANVOSSLNAAATATAVSDSASEVASASESA

ANSSASAPSNSSAISEVAPSSYGNSTIAQPSTSTKSDASI TGPITTDKVI TN

ESGIVFTSTVILTHVSEYCDQTSAAAVQSSACEEQSSAKSEQASASSSEQVKJ1T5VVM
CESSIQIESVKTSAEAAHKTEVIASCASELSSLSAKSEAMKTVSSLVEYQSAVAK

Q7SLAAVQSSAA5VQL5AAHAQKSS5EAVEVAQTAVAA5K5K5DEISTEIVNI1TKTVSS
GKETGV5QATVAANTH5VAIANMANTK5FASTMSLLV5F5FVGLF1"

	COUNT		COUNT		COUNT
IN	1327	a	1009	c	932
			g		1524
			t		

Item Scores:

Length:	4792
Matches:	49
165.50	

Percent Similarity:	38.69%	Conservative:	28
Local Similarity:	24.62%	Mismatches:	89

Indels:	33
Gaps:	6
Match:	16.18%
	8

0
-081-935-4 (1-194) Y AE25A147 (1-4793)

14 C[udoub]aGera]aG[nt]auBröaG[ol]aBröG[us]a]a]auAanG[us]p[he]va]G[] 33

[illegible]

34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1

[illegible]

1376 AACACTGGTGTGACCCCAATGTCGA---TAC TGGATAC TGGGCGT TGTGTAT TATGCCA 1432

54 GIuleuProGIyGlnIleIhrProCysAlaGIuGlnAlaCysProLeuAspAlaArgIle /3

1433 ACATTGCTGGTCCATTGGTCTTGTAATGCTGAGAAAGTAAAGCCAAAGCGTTGTT 1492

```

74 SerValSerAsnIleValValAspGlnCysSerLysAlaGly----- 87
||| :||| ||| |||
||| :||| ||| |||

```

1493 TCTGCTACAGTTTGGGAATTCCATTGTTCGGTGTGCTGTGGATCCATACTGG 1552

```

88 ---ValProfileaspile----- 92
      |||||
      : : : :

```

1553 ATGGTGCCTGCAATATGTCACAGACAGATTAACTGCTGCTGCCACTGCTGTTGCATCGCTT 1612

```
93 -----ProValaAspThrThrAlaAlaProGluProSerGluThrAlaGluProThr 110
```

1613 TCTGAACAACAGTTGAAACATCTTCTGAACACAGCTGATCTTCTCAGTCTGTTGAATCT 1672

111 AlaGluProThrGluGluProThrAlaGluProThrAlaGluProThrAlaGluProThr 130
.....

1673 TCTCAACCTGCTGAACCTCATCTGAACCTGCTGAGACTTCATCATCTGAACCTGCT 1732

Oy		131	HiegiuProthngluHiegiuProthrAlvalProthnglytnglyGlylvalProthr	150	
				:::	
Db		1733	GAGACTTCATCGAACAACATCATCCGAACAACCTGCTTATCTGAACCTCGTAAACTTCA	1792	
Oy		151	GlythrtglsrPheThvAlthrtglyrgrProthralAserrthPrroalglu-----	168	
				:	
Db		1793	TCAGAAGATCTTTCTACATCATCTTGACCACCCA-----TCACCTCGAAGATTACCCA	1846	
Oy		169	-----PheprroGlyAlaglysernsnValAglaAserrValgylleAla	184	
				:::	
Db		1847	TACACCATCTACCCCAAGTGTGCCAAGACT-----GCTTATCATATGGTTTTGCT	1897	
RESULT 23					
AC095003/c					
LOCUS					
DEFINITION	AC095003	238046 bp	DNA	linear HTG 09-MAY-2003	
VERSION	AC095003				
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	Rattus norvegicus (Norway rat)				
ORGANISM	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
	Rattus.				
REFERENCE	1 (bases 1 to 238046)				
AUTHORS	Munuy,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Albrooks,S., Amin,A., Angiano,D., Anyalabechi,V., Aoyagi,I., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barneslead,M., Behamed,F., Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Butrell,K., Calderon,E., Cardenas-V., Carter,K., Cavazos,I., Cesaar,H., Center,A., Chacko,J.J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Delgado,M.L., Davis,C., Davy-Carroll,L., De Ande,C., Dedetich,D., Diaper,O., Denson,S., Detamo,C., Ding,Y., Dinh,H., Divya,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Gantta,R., Garcia,A., Garner,T., Garza,M., Gegeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulik,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Khan,Z., King,L., Kovac,C., Kwis,C., Kraft,C.L., Lebow,H., Lavan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lotensheha,U., Louiseged,H., Lozador,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindratne,M., Mahmood,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Manthey,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,K., Mudida,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Paetzeltemuh,O., Okwonu,G., Olampunsaagon,A., Pal,S., Parks,K., Paternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C., Plopper,F., Polndexter,A., Popovic,D., Primus,E., Pu,L.-L., Pluzo,M., Quiroz,J., Rachlin,E., Reeves,K., Reiter,M.A., Reigh,R., Rellly,B., Rellly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Saverly,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Sheely,J., Sivatsreyen,A., Slason,I., Slinger,C.D., Snajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steidle,M., Strong,R., Sutton,A., Sytek,A., Taboc,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Uemali,K., Valas,R., Vera,V., Villasan,a.D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von				

```

TITLE      Medenhausen,A., Weis,R., Smith,D.R., Holt,R.A., Smith,H.O.,
JOURNAL    Weinstein,G. and Gibbs,R.A.
REFERENCE  Direct Submission
AUTHORS     2 (bases 1 to 238046)
TITLE       Worley,K.C.
JOURNAL     Direct Submission
            Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            3 (bases 1 to 238046)
REFERENCE   Rat Genome Sequencing Consortium.
AUTHORS     Direct Submission
JOURNAL     Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA

COMMENT----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GBH
Center clone name: CH230-6K12
----- Summary Statistics
Assembly program: Atlas;
Consensus quality: 226044 bases at least Q40
Consensus quality: 227924 bases at least Q30
Consensus quality: 229373 bases at least Q20
Estimated insert size: 234613; sum-of-contigs estimation
Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 236832: contig of 236832 bp in length
* 236833 236932: gap of unknown length
* 236933 238046: contig of 1114 bp in length.
Location/Qualifiers
1. 238046
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-6K12"
1. 1680
/note="wga_end_extension
clone_end:T"
6066_6753
/note="clone_boundary
clone_end:T"
site:BCORI
end_sequence:BH360832"
misc_feature
228068 . 228893
misc feature

```

```

/note="clone boundary
clone_end:Sp6
site:ECORI
end_sequence:BH360837"
misc_feature
229310..230352
/note="wgs_end_extension
clone_end:Sp6
235657..236832
misc_feature
/note="wgs_end_extension
clone_end:Sp6"

BASE COUNT 61790 a 49774 c 51497 g 67565 t 7420 others
ORIGIN

Alignment Scores:
Pred. No.: 23.1 Length: 238046
Score: 165.50 Matches: 32
Percent Similarity: 58.02% Conservative: 15
Best Local Similarity: 39.51% Mismatches: 31
Query Match: 16.18% Indels: 3
DB: 2 Gaps: 1

US-10-081-935-4 (1-194) x AC095003 (1-238046)

QY 93 ProProvalasprThrrAlaAlaProGluProserGluThrrAlaGluProThrrAlaGlu 112
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167953 CCACCTACCAAGCCACTAGCCACCAACCACTCCACCAACCAACCAAGCCACCA 167894

QY 113 ProThrgluGluProThrrAlaGluProThrrAlaGluProThrrAlaGluProThrrAlaGlu 132
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167893 CCACCAACCAAGCCACCAAGCCACCAACCAACCAACCAACCAACCAACCAACCA 167834

QY 133 ProThrgluGluProThrrAlaAlaProThrgluGluProThrgluGluProThrrAlaGlu 152
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167833 CCACCAAGCCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 167774

QY 153 GlyserTherhrrValThrrGlyArgProThr-----AlaserThrrProAlaGluPhe 169
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167773 CCACCAACCAAGCCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 167714

QY 170 Pro 170
|||||
167713 CCA 167711

RESULT 24
AC098905 225849 bp DNA linear HTG 10-MAY-2003
LOCUS Rattus norvegicus clone CH230-163J19, WORKING DRAFT SEQUENCE, 2
DEFINITION
unordered pieces.
ACCESSION AC098905 GI:30521215
VERSION AC098905.6
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULUTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 225849)
Muzny,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alsbrooks,S, Amn,A, Anguiano,D,
Anylenbech,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Blawie,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryan,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
Cadenas,V, Carter,K, Cavazos,I, Cesar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleaveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Davis,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dedrich,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
Diaper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M,
Gebregiorgis,E, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,W,

```

```

Gunaratne,P, Haaland,M, Hamil,C, Hamilton,C, Hamilton,K,
Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hoque,M,
Hollins,B, Howells,S, Huyl,S, Hume,J, Idlebird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kover,C,
Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,J, Liu,M, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorenznewa,L, Louieged,H, Lozano,R,J, Lu,X, Ma,J,
Maheshwari,M, Mahindartine,M, Mahmoud,M, Malloy,K, Mangum,A,
Mangum,B, Mapa,P, Martin,K, Martin,R, Martinez,E,
Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Mondasa,M, Murphy,J, Moore,S,
Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Parks,K,
Nwaokelenen,O, Okwunu,G, Olarinpinuagoo,A, Pal,S, Parks,K,
Pasternak,S, Paul,H, Perez,A, Perez,L, Plannoch,C,
Plopper,F, Polindexter,A, Popovic,D, Primus,E, Pu,L,
Puzo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J,
Sanders,M, Saverly,G, Scherer,S, Scott,G, Shatsman,S, Shen,H,
Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajs,D,
Sneed,A, Sodergren,E, Song,X,Z, Sorelle,R, Sosa,J,
Steinle,M, Strong,R, Sutton,A, Svatek,A, Tabor,P, Taylor,C,
Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K,
Vaisas,R, Vera,V, Villaseana,D, Waldron,L, Walker,B, Wang,J,
Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F,
Williams,G, Willson,R, Wleczky,R, Wooden,H, Worley,K,
Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V,
Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von
Niederhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O,
Weinstock,G, and Gibbs,R,A.
Direct Submission
Unpublished
2 (bases 1 to 225849)
REFERENCE
Mozley,K,C.
Direct Submission
Submitted (05-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 225849)
REFERENCE
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:2285625.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented. The sequence
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GIOC
Center clone name: CH230-163J19
----- Summary Statistics
Assembly program: Atlas 3.0:
Consensus quality: 212145 bases at least Q40
Consensus quality: 214496 bases at least Q30
Consensus quality: 216260 bases at least Q20

```



```
molD). inositol 5-phosphatase 2"
/protein_id="AA052441.1"
/db_xref="GI:28829951"
/translation="MCGNKKHOACIQCAVRSKDFEGLCEDNCHTVEGFTQEV
EPEENGCKYKTSYKRVWTOEVLDAWKKLRKDKRYYILQWLKIQVTSYSLKIDGY
NQOLYOVTELMFKLIDKFSQDEGSSSTSSSTSTSTFTGIIINYPPEITLSCYLA
CYLLRPVITISDFISLAQSGKIASFTNETTWSMNSSTFNFARISDDICKYNAL
LSSSNPLLLRFCELTCLKPKSLINLNOLYNLILPKLFRKSPNLSLTKIYENIL
LPNSDILCVSIIIMILKSKFNLFDROYNPSSGSSNNNNNTTLDNNNNNNNEI
NNEYDNNIDKNYKENSIERLEFLKLIINNPNSENSENSISSEERIKEEYKRNEL
ENWLGVYINLPKNEFLSKODTWSLIDQFKGIPLKPSISIRLITDQHSKYSLSID
KYQNLNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
SKSKSSLSLSSSSSSSSADPTLLPKVNRINRAKINSRGRTNLPPEQKSVFSKGT
GKRG"
6218..9748
/note="GeneID exon scores (in order of location ranges) :
356..10 - GSCJ_ID dd_03095"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). kinase responsive to stress 1-like kinase"
/protein_id="AA052442.1"
/db_xref="GI:28829952"
/translation="MSNLKLSNNNGNOKESSFFQKVKSPSTONLNSFSSNNNS
NLSNGSNEVKDITNSPQLPENYTPPPHQRSSSIEGGEFSLNNNSDNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
TDDSTTTEITVSGMSPQOLNSNIDSSQOHODLSSVNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
PGLISPRSDNOLPEKKKKKKFKTPPEIKFHHHKKSSLSSTTTPETTSISPS
SSSLAISPPNTAATTNNKSKHKTKSTFDINTELISVYNIHKHVPFDLKTGHN
DTLDEKLDGAGVYKGTHTDGLFTLAIKVIEMKESVSLOEINILKQKSPNI
VSYFSLQESHAWILDFECALGIRDIIESTEKLENAQISFVVKNTLKGILYLSQ
NIHRDVKANVLESCDVKIADFGVSEKNGALDOSEKIMIGTPLMAAPKVIILKYS
DYKADIMSLGITIEMADGLPHILDPKRAMKWNMPPEPKKMSPELLNKYLA
RLVLDPEKRAPIIDLCHPLKDRGPRVLSLVNQLKIKKKIKDILKKQKHDS
QSSSSSPSPNATVNGDIDGSLSTSIISPPSPDLNCSNLSSVLSKML
SKRGVYTRDRDQEEEDSKLNNDQEDQDDEDDDEDDDEDDVDFSTIIFHG
KKKSGGNGGAGVTSDDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
NKTSTKSNKKKKKNNLSTIGKSGGNLHVAHPKPLPTTPPFSISMNEKOLETK
LETTYIDSNQIKVINDIKNEIKOLESIINKINLOQOLPILALEEKONHDTX
PROMOSKLSATNLNEKILSSSPSSPLTNSVNSLITTTTTPTSVPULSRGSSRRS
GSISSSSSPSSPLSSSDGPAFNSNSTTTTNSSSSSSGGSGVDSIPMTGASPSIM
KRTTSSSSSPSSPLSSSDGPAFNSNSTTTTNSSSSSSGGSGVDSIPMTGASPSIM
NSSGSGNSPSLSTSSSTNSNSVNNIK"
join(10002..10168,10786..12082)
/note="GeneID exon scores (in order of location ranges) :
0..26, 60..31 - GSCJ_ID dd_00942"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA052443.1"
/db_xref="GI:28829953"
/translation="MESIFSYLNNYDCGNDEEFLNCLLILSINNLITFENLLY
CYNNLKRKTLNFKVNLIIEDNNGETPLNRPYPLPYGVREVDSSSTTG
DHWIPDTCKSYFHNPNQOKLNRIINSLTTFIFGNSNOPLSSNGIPWPKLKT
LSLQSGFOOTIRNELPQSLTEPLDSEIQIILNESIPKSVYTLTAYEFTIRKHPD
TKSISKGVSTLKGICYFNQITKDMIPIAVTSLIEDNISKEIENPSLSTIRKHPD
SIELDSVILPENIKLSIYGDSHSCFLPKSIESKITSFNLKISFNPNLSTSLK
IEKLSDSVILEGVPNTLRLIVLSCPNLKSNGIKPNSLKIKLKEPNPIANGPLPS
SIEVLKPNVYDPIKNGDPLSPLVNLKODHCTISITIODALPNNKLKPYFHRELRKL
PNLWPTSPPEFVIGSEASFPNFSIDLDFETHIKLK"
join(11253..12311,12539..13007)
/note="GeneID exon scores (in order of location ranges) :
-5..63, 15..15 - GSCJ_ID dd_00941"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA052444.1"
/db_xref="GI:28829954"
/translation="MDNITITSLINNYYNNKNEEKEEKHLNDLKLNDQOFTLN
NHIIQNLKNTFTTNSNDVINYHNSITDNNNSNDIINNKKIADCSILLIINKNLN
HHIITKNNNSNNLNLILKPPDLNKLRTQOQIDQITDNNYKLFKVMGNIIFRDO
ILFLRLIYNNHNSI"
join(13144..14052,15655..15768)
/note="GeneID exon scores (in order of location ranges) :
36..56, 7..78 - GSCJ_ID dd_00940"
```

```
/codon_start=1
/product="hypothetical protein"
/protein_id="AA052445.1"
/db_xref="GI:28829955"
/translation="MRCSDIILVDHNIPTCKSIIFSHNQOLTRNIIIGSLTSDPG
KPNQSLSDKNGPMLPRLKLSIFPKSGQHSIHREIPSLSLIDPYKKEVIEG
SIPKSTVTLAHYVDSKGAKLDSITPGGTSTLSEFPFOOTLEKMISSNNTVKS
FNSGISKSLSPISITLHLSGNSQNPPOVPOATNLSYTTSSDIKILPKTIQKLI
MPSPPSQLDNHSFYLKULSTLNFDAKQVDLNVFPIINSLTLAVSDYLEFNOL
LQNFELKLLFNFNQPLNVGDLPSISIEGCPAYVAIRANGLIQALLSDHGLDVNNLIT
FNQLOD"
join(15826..16011,16072..17106)
/note="GeneID exon scores (in order of location ranges) :
1..92, 31..02 - GSCJ_ID dd_00422"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA052446.1"
/db_xref="GI:28829956"
/translation="MTQLTRMKRIAFDCSILIIKNNLNHHITKNNLNHTTNSNN
KNNLNQVILKLIQINKGEIYLSQIKYFLEYIITIIISIDIGGHRISTYKRN
YRNSVINSDENNHQONISNCEPMPSTAEKIEKMSDIIIVDHWIPDTCKIKTY
FDQQLTRNIIIGSLTISINFSNEFQSLDSTGEPMPLPRLKSLTFKSGQHSIHREI
PSLTLTLEFIPYKGVIOYGSIPKSVTLTLYSLFFKEHDIYKFDLTSVPGTSLSEF
FYNGTIEKGMIPHYVTSLEKFRMGDKLEPNSLPISIKTLGCRNRTIFQKVPPTI
TULSITVTFSDIGILPXTIEKLIIGSNFTQDLSHSFHLFENKLSLMAIGVDLTGV
ESPRTIKTLTLVL"
join(18782..18911,19094..20841)
/note="GeneID exon scores (in order of location ranges) :
0..11, 70..39 - GSCJ_ID dd_00420"
/pseudo
/codon_start=1
join(21335..21528,21616..21928)
/note="GeneID exon scores (in order of location ranges) :
-5..52, 10..92 - GSCJ_ID dd_00435"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA052447.1"
/db_xref="GI:28829957"
/translation="MDSITFYLINNNYNNKNEELNCLLILSINNLITFQVODYC
NSINNYKNYONLQKKEKLEQEEBEELNDLKGIDQOFTLNHHIIONLQKFTTN
NSNDEITYHSGITDNNNDTIDNRKRIAFDCSILIIKNNLNHHITINNKKINN
NLQILK"
complete(22154..22339)
/note="GeneID exon scores (in order of location ranges) :
0..40 - GSCJ_ID dd_00433"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA052448.1"

Alignment Scores:
Pred. No.: 14.8 Length: 110000
Score: 163.50 Matches: 54
Percent Similarity: 37.71% Conservative: 12
Best Local Similarity: 30.86% Mismatches: 82
Query Match: 15.98% Indels: 27
DB: 3 Gaps: 4

US-10-081-935-4 (1-194) x AC116957_0 (1-110000)
QY 22 AAPILEPPOPCyAlaLeuabnCySpheValGluAlaLeuGlyAAsnAP----- 38
DB 48800 GATGTTCCAGTATTATACAGGTTCTATAGAAATTTCTACTTAGGTCGATTTATGCTTA 48859
QY 39 -----GlyCyethrArgLeuThrAparPhelyCyAHisCyserIySpGoluLeu 55
DB 48860 ACAGSTAAAGATTAAATCACTACTGTCGACATTCATTAATATCACTACTAA 48919
QY 56 ProGlyGlnIethrProCyValGluGluAlaCySpProLeu----- 69
DB 48920 CCATTACAAAGTAAAAAAGCTGTGGGCGTCTGCTATTATTTAGCGCGATCTCTGA 48979
QY 70 -----AspAlaArgIeserValSer 76
```

```

Db      48980 ATGGATCAACAGTACTTACTTATTTATTTAGAAAATGATAAATCAATATCTTCAAT 49039
Oy      77 AsnIleValaIaspInCySeSerLyAlaGlyValProIleaspIleProValaIasp 96
Db      49040 GAATATTTATTTATTCATATCAAAAAGTATTTATTTATTTATTCAGATCCACCAACCCCA 49099
Oy      97 Thr---ThraIAlaIaProGluProSeGluThrAlaGluProThrAlaGluProThrglu 115
Db      49100 ACTCAACACCAACCAACCACTGAAACCAACCAACCAACCACTCAACCAACCACTCAAA 49159
Oy      116 GluProThraIaIaIaProThraIaGluProThraIaGluProThraIaGluProThrglu 135
Db      49160 ACACCACTCAACACCACTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAAA 49219
Oy      136 GluProThraIaIaIaProThrglyThrglyGlyValProThrglyThrglySerPhe 155
Db      49220 ACACCACTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCA 49279
Oy      156 ThrValThrglyThrglyProThraIaSerThraIaGluProThrglyThrglySerPhe 170
Db      49280 ACACCAACCAACCAACCACTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCA 49321

RESULT 27
AC119773 222896 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-150P22, *** SEQUENCING IN PROGRESS
DEFINITION
AC119773
AC119773.6 GI:25095179
VERSION HTG: HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus
            Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
            1 (bases 1 to 222896)
            Musny,D.Marie, Metzker,M.Lee, Abramson,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Bismawo,K., Blair,D., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Devilla,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,M.,
            Gunaratne,P., Haaland,M., Hamill,C., Hamilton,C., Hamilton,K.,
            Haver,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Huliy,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Joliver,A.,
            Kapachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowitz,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Loresenheva,L., Louissege,H., Lozada,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapue,P., Martin,K., Martin,R., Martinez,E.,
            Mawhinley,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Miliostavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Mundaasa,M., Murphy,M., Natr,L.,
            Nankervis,C., Neal,D., Newton,N., Ngundea,N., Norris,S.,
            Nwokilemeh,O., Okwomu,G., Olampunsgoon,A., Pal,S., Parke,K.,
            Paternak,S., Paul,H., Perez,A., Popovic,D., Primus,E., Pu,L.L.,
            Plopper,F., Poindecker,A., Popovic,D., Primus,E., Pu,L.L.,
            Plazzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,R., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Sander,M., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Soer,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umanli,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wenden,H., Morley,K.,
Williams,G., Willison,R., Wleczky,R., Wooden,H., Morley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G., and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 222896)
Morley,K.C.
Direct Submission
Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 222896)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:22856367.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVPs
Center clone name: CH230-150P22
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 206295 bases at least Q40
Consensus quality: 208887 bases at least Q20
Estimated insert size: 209030; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/centbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 7 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
130230: contig of 130230 bp in length
130231 130330: gap of unknown length
130331 150793: contig of 20463 bp in length
150793 150893: gap of unknown length
150894 217180: contig of 66287 bp in length
217180 217280: gap of unknown length
217281 218456: contig of 1176 bp in length
218457 218556: gap of unknown length

```



```
repeat_region 1759. 1839
repeat_region 2318. 2683
CDS            3625. 4284
              4253. 4253
```

```

/codon_start=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAB54077.1"
/db_xref="GI:584339"
/translation="MKKPKKIIINOICNLVKKLNDDDCGFKXEYQODPFDDLQSFNF
GLFSSKINNNQNDHNENHNAQDNNNSRKIIASFPLKNGHSISLKULDLNLSR
IGLKFVQGLTQGLIDIVYEKKMGNGVNFIEHKVDYCFDMIDEDLIQHLFSDNNR
NNNNEDEDLDDLDNNNSSEKLSWMSKINSLYVOIYPKLSYKIKESIGNKXIKDKIQ

```

BASE COUNT	1767 a	615 c	432 g	1470 f
ORIGIN				

Alignment Scores:	
Pred. No.:	0.901
Score:	162.00
Percent Similarity:	43.364
Best Local Similarity:	34.274
Query Match:	15.844
DB:	3
Length:	428
Matches:	49
Conservative:	13
Mismatches:	55
Indels:	26
Gaps:	6

US-10-081-935-4 (1-194) x DDU20661 (1-4284)

53 ProGluLeu-----ProGlyGlnIleThrProCysValGluGluAlaCysPro 68

Db 2152 CCAGATTAAATCAACATCAAGTCTTCAATATCTGAATTTTCTGAATCTTGCAACCA 2211

Qy 69 LeuAspAlaArgIleSerValSerAsnIleValValAspGlnCysSerLys----- 85

Db 2212 TTGGATCTTCTGAATAATTCTCAATCATTGACATCTTCTGAATCTTGGCAATCATTTGGA 2271

Qy 86 -----AlaGlyValProIleaspIlePro-----Pro 94

Db 2272 TCTTCTGATCTTCACAAACCCTTCGTCACTCCTGAAGATCACCAACCACTCCA 2331

QY 95 ValAspThr-----ThrAlaAlaProGluProSerGluThrAlaGluPro 10

Db 2332 ACTGAACAGAAACACCACTCCAACTGAAACGCCCACTCCAACTCCA 23391

QY 110 -----ThrAlaGluProThrGluGluProThrAlaGluProThrAla 111

Db 2392 ACTGAACGCCCACTCCAACCTGAACACACCAACTGAACACCAACTGAA 2451

QY 128 GIUProThrHisGIUProThrGIUGIUProThrAlaValProThrGIUThrGIUGlycyl 14

Db 2452 ACACCAACACCAACCGACTGAAACCCCAACTGAAACACCAACTGAAACACCAACGCCA 2511

148 valProthrglyThrglySerphethrValThrglyargprothraAlaserThrproAla 10

Db 2512 ACACCAACGGAACTCCCAACAGAAACCCCAACAGAAACCCCAACA--GAAACCCCAACA 2560

QY	168	GluphePro	170

Db 2569 GAAACCCCA 257

RESULT 29			
AC095211/c			
LOCUS	252801 bp	DNA	linear
AC095211			HTG 03-OCT-2001
Rattus norvegicus clone CH230-9E3, ***		SEQUENCING IN PROGRESS ***	
DEFINITION			

REFERENCE	3 (bases 1 to 252801)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (03-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Sep 10, 2002 this sequence version replaced gi:2172802.

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebech,I., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Badarinarake,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Butch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Chen,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Hochs,S., Dunn,A., Durbini,K., Duval,B., Evans,K., Egan,A., Escotto-Mo, Eugene,C., Evans,C., Fallis,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Fraser,C.M., Gabits,I., Gante,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Geer,K., Gill,R., Gaddy,M., Guerra,M., Guevara,M., Gunatratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladin,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Huliy,S., Hume,J., Idlebit,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowic,C., Kraft,C., Lebow,H., Levin,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,M., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenzuela,L., Louisedge,H., Lozada,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mahaney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naif,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokwelench,O., Okwonu,G., Olarinmusaogun,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plummer,F., Polidexter,A., Popovic,D., Primus,E., Pu,L., Prazo,M., Quiroz,J., Rachin,I., Reeves,K., Reiser,M., Riggs,F., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rivers,C., Rodkey,T., Rojia,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,M., Saevry,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajz,D., Sneed,A., Sodegiren,E., Song,X.-Z., Sorelle,R., Sooa,J., Steidle,M., Strong,R., Sutton,A., Swalek,A., Taboi,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villaasna,D., Waldron,L., Walker,B., Wang,J., Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willison,R., Wleciyzk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,Y., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausen,A., Weiser,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 252801)
Worley,K.C.

Direct Submission
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 252801)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (03-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 10, 2002 this sequence version replaced gi:21722802.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atlas/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

-----Genome Center
Center: Baylor College of Medicine
Center code: BCM

gene /db_xref="GI:28056731"
/translation="MRSAFVFPYFILRPVAVMNNRSITCSPAQNRQAGMSLEI11
VLLIGVAVFGSRVLAGDAGKALVKSQIOTLAGKIENYOLDGKLPGLNDLVN
PFGVSNMVGPAKAAELNDPMGHPIQYRVPEGRPFDMISLGKDGKPGSSSYDADIK
VE"
607..1113
/gene="xpsH"
/locus_tag="PD0735"
607..1113
/gene="xpsH"
/locus_tag="PD0735"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="general secretory pathway protein H precursor"
/protein_id="AA028604.1"
/db_xref="GI:28056732"
/translation="MLPCSAFPFRFQSSGAPVPDAQGVSLLEMLVIALIALISLGLVA
VLSGMDVQVLRNSKQOLAELRYTRVALATRAQRFMLDPQRRWQGNHOGKLP
GALQIRFSGAQOQLQTSSEGGVIFQFDDGASSGGRIDLEMKRAYRIDVGWITGEVRS
PLRTSASL"
1110..1508
/gene="xpsI"
/locus_tag="PD0736"
/note="synonym: pefI"
1110..1508
/gene="xpsI"
/locus_tag="PD0736"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="general secretory pathway protein I precursor"
/protein_id="AA028605.1"
/db_xref="GI:28056733"
/translation="MKCGRGYSLELVAVAFAILALSLILGTLGTGTPRVRADLAS
OATLVASFLAQCGVAMPRLHVQOQGFEGFRRMLETRPRENSPSPGAOLLELI
LSVRGDEPQOSIQMRTLRFVRLQPPDGL"
1505..2131
/gene="xpsJ"
/locus_tag="PD0737"
/note="synonym: pefJ"
1505..2131
/gene="xpsJ"
/locus_tag="PD0737"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="general secretory pathway protein J precursor"
/protein_id="AA028606.1"
/db_xref="GI:28056734"
/translation="MRMAGFTLMEVLLATVLLASGLTFAFTLHSASIVSRSEALA
SRERRAVELFLRRMGALPLMGGEDRRHMLFVGEPQWRPFADVPDLVGRGC
PYHDISVSGVGHQRLALTLHQSDETKDGLSVSELLADGLHEVHSYKCMDS
RHQDGPVATWPSDDQLPVSVAQIKDIDAPWPLVLLVLLQNSDGD"
2131..2976
/gene="pefK"
/locus_tag="PD0738"
2131..2976
/gene="pefK"
/locus_tag="PD0738"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="general secretory pathway protein K"
/protein_id="AA028607.1"
/db_xref="GI:28056735"
/translation="MRRAGALIVLMLVLAAMIGAFALSAVHESLQAGVLLVSSA
QAOEVNAGLEVAVSRLODSVQVRWIPQGRPYSMYADATLEMRMIDSGKVDLVV
ADQLLTALIQALGVDAVARAARLAGAIIWMRDTHLVSPGAEBSDYAAAGLPYGAGD

gene NPFQSIARELVLGMDTDLYRLMLPNIITYSNLOPEVSAPVPLTALGVNAVLA
KPOQSPSDVSVKQMSGTTSTESKRLACGAEVLAAYVRMGGRIPGATYITLHMBE
GIAMQ"
2949..4103
/gene="pefL"
/locus_tag="PD0739"
2949..4103
/gene="pefL"
/locus_tag="PD0739"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="general secretory pathway protein L"
/protein_id="AA028608.1"
/db_xref="GI:28056736"
/translation="MGRGYRMAKMSKMSLRFNGVSIABQVGFMMHMQSLLAUPLM
RCRVOMGLSERLLSVQADGHLVYQCDVDEPLVQLPMPITPOELSGMLPKQLTL
PRHWLIPATSAICRPLRLPAGASORLHDVRFEDOTPEQVYVDVRLKRGDD
AQETELVYVPRRLVNGPQVREKMTTLAGIDVAMOGQPLNVLPLPVQRWCSDP
MRWNILAAVALLVATGALLNRRIVABQLRQVQKQASARVAERORLVTL
IEGVRLDERRTLPRVIEVWELSRLLPNGTYLERFALIEDGOLHMGNSDSSLVR
LLEGSKLWRTPSLSVLSQNVGSSVNRVFIETALVGVPWKEADGKAOR"
4084..4716
/gene="xpsM"
/locus_tag="PD0740"
4084..4716
/gene="xpsM"
/locus_tag="PD0740"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="general secretion pathway protein"
/protein_id="AA028609.1"
/db_xref="GI:28056737"
/translation="MARHVNSDRMALGLFIKLSVLYLLVHPMFTPRMYSVQOQI
OVRDELEIITKQLOCGANVLEQRLRIQVLRPGRLPSSVLSAGLVKRLGVV
NDSPGNRGCATITNRSPLQDMDRFRVAVQVRLCGTEFLALVYNLESGRPLFV
NMLNVAQRYPPLSPNSGCHGLDVAFELAGFLSPGCHVPSLPORSSNAN"
4706..5473
/gene="xpsN"
/locus_tag="PD0741"
4706..5473
/gene="xpsN"
/locus_tag="PD0741"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="XpsN protein"
/protein_id="AA028610.1"
/db_xref="GI:28056738"
/translation="MRIDAFSVWTWLFATVAGMAALVCMALAGMGRSVOLRBDSQL
IRKQLPASPAAHEGPFTRGELIARPVFSRDRKHPPLVSGEITSVEDRLTG
VLTTRPKATYLTNEONOSLRIGDQAVAVQVRLCGTEFLALVYNLESGRPLFV
NGDANKVANNPSENVPTRVEHVRVAGSSSAAALVLPKSTSAVSQSGRPGCGQPAV
VTGPQAPKXPIDAQIRAIRERIREARERLOOQOEOQK"
5518..7845

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
48.9	300029	39	4	49	1	1
Score:	161.50					
Percent Similarity:	46.24%					
Best Local Similarity:	41.94%					
Query Match:	15.79%					

US-10-081-935-4 (1-194) x AE012556 (1-300029)

Qy 79 ValValAspGlnGysSerLysAlaGlyValProIleAspIleProProValAspThrThr 98
|||:::|||||
Db 252627 GTCTTAGAGCTCCTGTGCTGATTAAGTTCGATAGATACACCTACACGACACCCACA 252568

[illegible][illegible]

Db 1124 TTTTATT 1116

RESULT 32
AC109525 0/c

WPCOMMENT

Sequence split into 5 fragments LOCUS AC109525 Accession AC109525

Fragment Name	Begin	End
AC109525_0	1	110000
AC109525_1	100001	210000
AC109525_2	200001	310000
AC109525_3	300001	410000
AC109525_4	400001	478059

LOCUS AC109525 478059 bp DNA linear HTG 11-OCT-2002
DEFINITION Rattus norvegicus clone CH230-255D23, *** SEQUENCING IN PROGRESS
*** 5 unordered pieces.

ACCESSION AC109525 GI:23603814
VERSION AC109525.6
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 478059)
Muzny,D,Marle, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Albrooks,S, Amin,A, Angiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Bisswaio,K, Blair,J, Blankenburg,K, Bluth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gburegogis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W, Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hughes,M, Hollins,B, Howells,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,Z, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenzowa,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindartne,M, Mahmoud,M, Malloy,K, Mangum,A, Mangun,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mahoney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,S, Nankevis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwokolemele,O, Okunou,G, Oiarunaggon,A, Pal,S, Parks,K, Patelnak,S, Paul,H, Perez,A, Perez,L, Pflannkuch,C, Plappert,F, Poindexter,A, Popovic,D, Primus,E, Pu,L,L, Puzo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S, Sanders,W, Savery,G, Scherer,S, Scott,G, Shatman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smajls,D, Sneed,J, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J, Steimle,M, Strong,R, Sutton,A, Svatek,A, Taber,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K, Valas,R, Vera,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,O, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wleczyk,R, Wooden,H, Worley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,Z, Zhao,S, Dunn,D, von Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 478059)
Morley,K.C.
Direct Submission
Submitted (05-PEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 478059)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Oct 9, 2002 this sequence version replaced gi:22795157.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPH2
Center clone name: CH230-255D23
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 380245 bases at least Q40
Consensus quality: 385911 bases at least Q30
Consensus quality: 389757 bases at least Q20
Estimated insert size: 390792; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
NOTE: This sequence may represent more than one clone.
NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

Location/Qualifiers	1	235023	contig of 235023 bp in length
235024	235123	gap of unknown length	
235124	246065	contig of 10942 bp in length	
246066	246165	gap of unknown length	
246166	474601	contig of 228436 bp in length	
474602	474701	gap of unknown length	
474702	475761	contig of 1060 bp in length	
475762	475861	gap of unknown length	
475862	478059	contig of 2198 bp in length.	

FEATURES

source

1. 478059

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-255D23"

1. 2769

/note="wgsc_end_extension"

clone_end:5p6"

misc_feature

```

misc_feature      complement(9835..10732)
                  /note="clone boundary
                  clone_end:Sp6
                  site:Tbcl
                  end_sequence:RXACK24TVB"
misc_feature      235124..236699
                  /note="wgs contig"
misc_feature      287702..289205
                  /note="wgs contig"
misc_feature      350281..351366
                  /note="wgs contig"
misc_feature      358886..361234
                  /note="wgs contig"
misc_feature      406035..407045
                  /note="wgs contig"
BASE COUNT      110790 a 84236 c 86247 g 111645 t 85141 others
ORIGIN
Alignment Scores:
Pred. No.:      22.1      Length:      110000
Score:          160.50    Matches:      34
Percent Similarity: 52.27% Conservative: 12
Best Local Similarity: 38.64% Mismatches: 39
Query Match:    15.69% Indels: 3
DB:             2        Gaps: 1
US-10-081-935-4 (1-194) x AC109525_0 (1-110000)
QY      86 A1AG1yValProT1eAsp1eProP1oValAsp1eTh1A1a1aProG1uProSerg1u 105
DB      53334 GCCAGCCAGCAGTCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 53275
QY      106 Thr1AgluProTh1AgluProTh1AgluProTh1AgluProTh1AgluProTh1AgluPro 125
DB      53274 ACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 53215
QY      126 Thr1AgluProTh1AgluProTh1AgluProTh1AgluProTh1AgluProTh1Aglu 145
DB      53214 ACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 53155
QY      146 GlY1G1yValProTh1G1yTh1G1ySerP1eTh1ValTh1G1yArgProTh1----- 162
DB      53154 GCCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 53095
QY      163 A1AserTh1Pro1Aglu1uP1ePro 170
DB      53094 GCCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 53071
RESULT 33
AC094443 LOCUS      237019 bp      DNA      linear      HTG 09-MAY-2003
DEFINITION      3 unordered pieces.
ACCESSION      AC094443
VERSION      AC094443.6 GI:30466976
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 237019)
Muzny D, Marie J, Metzker M, Lee S, Abramson A, Adams C, Alder J,
Allen C, Allen H, Altschuler S, Amin A, Angiano D,
Aryal-ecchi V, Ayagi A, Ayodeji M, Baca E, Baden H,
Baldwin D, Bandaru A, Bandaru D, Barber M, Barnstead M, Benahmed F,
Bismail K, Blair J, Blankenburg K, Blyth P, Brown M,
Bryant N, Buhay C, Burch P, Buttrell K, Calderon E,
Cardenas V, Carter K, Cavazos I, Ceasar H, Center A,
Chacko J, Chavez D, Chen G, Chen R, Chen Y, Chen Z, Chu J,
Cleveland C, Cockrell R, Cox C, Coyle M, Cree A, D'Souza L,
Daya M, Davis C, Davy-Carroll L, De Anda C, Dederich D,
Delgado O, Denson S, Deramo C, Ding Y, Dinh H, Divya K,

```

```

Draper H, Dugan-Rocha S, Dunn A, Durbin K, Duval B, Eaves K,
Egan A, Ebecato M, Eugene C, Evans C, Falls T, Fan G,
Fernandez S, Finley M, Flagg N, Forbes L, Foster P,
Fraser C, Gabisi A, Ganta R, Garcia R, Garner T, Garza M,
Gebregorgis E, Geer K, Gill R, Grady M, Guerra M, Guvvara W,
Gunaratne P, Haaland M, Hamil C, Hamilton C, Hamilton K,
Harvey Y, Havlak P, Hawes A, Henderson N, Hernandez J,
Hernandez R, Hines S, Hladun S, Hodgson A, Hogue M,
Hollins B, Howells S, Huylk S, Hume J, Idebird D, Jackson A,
Jackson L, Jacob L, Jiang H, Johnson B, Johnson R, Jolivet A,
Karpathy S, Kelly S, Khan Z, King L, Kovar C,
Kowis C, Kraft C, Lebow H, Levan J, Lewis L, Li Z, Liu J,
Liu J, Liu W, Liu Y, London P, Longacre S, Lopez J,
Lorenzini W, Louieged H, Lozada R, Lu X, Ma J,
Maheshwari M, Mahindaratne M, Mahmoud M, Mallory K, Mangum A,
Mangum B, Mapa P, Martin K, Martin R, Martinez E,
Mawney S, McLeod M, McNeill T, Meenen E,
Milosavljevic A, Miner G, Minja E, Montemayor J, Moore S,
Morgan M, Morris K, Morris S, Muniasa M, Murphy M, Nair L,
Nankervis C, Neal D, Newton N, Nguyen N, Norris S,
Nwokeneme O, Okwuonu G, Olarinmusa A, Pal S, Parks K,
Pasternak S, Paul H, Perez A, Perez L, Plankoch C,
Plopper F, Polinder A, Popovic D, Primus E, Pu L,
Puzo M, Quiroz J, Rachlin E, Reeves K, Regier M, Reigh R,
Reilly B, Reilly M, Ren Y, Reuter M, Richards S, Riggs F,
Rivers C, Rodkey T, Rojas A, Rose M, Rose R, Ruiz S,
Sander M, Savery G, Scherer S, Scott G, Shatsman S, Shen H,
Shetty J, Shvartsbeyn A, Sisson I, Sitter C, Smajs D,
Sneed A, Sodergren E, Song X, Sorrell R, Sores J,
Steinle M, Strong R, Sutton A, Svatek A, Taber J, Taylor C,
Taylor T, Thomas N, Thomas S, Tingey A, Trejos Z, Usmani K,
Vale R, Vera V, Villaseca D, Waldron L, Walker B, Wang J,
Wang Q, Wang S, Warren J, Warren R, Wei X, White F,
Williams G, Wilson R, Wleczek R, Wooden H, Worley K,
Wright D, Wright R, Wu J, Yakub S, Yen J, Yoon L, Yoon V,
Yu F, Zhang J, Zhou X, Zhou X, Zhou D, von
Niederhausern A, Weiss R, Smith D, Holt R, Smith H, O.,
Weinstock G, and Gibbs R.A.
Direct Submission
Unpublished
2 (bases 1 to 237019)
Worley K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 237019)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:22772430.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GAP0
Center clone name: CH230-4F6

```


[illegible]

Pred. No.:	3..26	Length:	9636
Score:	158.00	Matches:	35
Percent Similarity:	46.02%	Conservative:	17
Best Local Similarity:	30.97%	Mismatches:	31
Query Match:	15..44%	Indels:	30
DB:	6	Gaps:	3

US-10-081-935-4 (1-194) x 195887 (1-9636)

OY	57	GlycylIleThrProCysValGluAlaCysProLeuAspAlaArgIleSerValSer	76
			::
Db	1515	GGATATTATTTACCTTGC-----	CTTCTCTTTTCA 1486
OY	77	AsnIleValValaPrgGlnCysSerLysAlaGlyValProIleAspIleProProValasp	96
		::::	
Db	1485	TCTACAAAT-----TCACCTTCTCTTACCCTTACCTCTTCCCTTCTTCTTCA	1438
OY	97	ThrThraIaaIaProGluProSerGluIuthraIagluProThraIagluProthrGluGlu	116
Db	1447	CCTACCTTTACCTACCTACCTTCTTACCTACCTCTTACCTTTCTTCTTACCTTCTTCA	1378
OY	117	ProThraIagluProthrIaIagluProThraIagluProThraIagluProthrIaIaglu	136
Db	1377	CCTACCTTGTACCTACTCTTCTTCAACCTTCTTACCTTCTTACCTTCTTCACTTCTTCA	1318
OY	137	ProThraIaValaProthrnglyThrGlyGlyValProthrnglyThrGlySerPheThr	156
			:::
Db	1317	CCTACCTTCTTACCTACTCTTCA-----	1294
OY	157	ValThrGlyArgProThraIaSerThrProIaGluPhe	169
			:::
Db	1293	-----CCTACCTTCTTCACTTCACTTCTTCACTTCTTCACTTCTTCACTTCTTCACTT	1267

RESULT 39
PPAS230A/c

LOCUS	PPAS230A	9654 bp	mRNA	linear	INV 29-DEC-1994
DEFINITION	Plasmodium falciparum (clone sv) s230 mRNA, complete cds.				
ACCESSION	L04162				
VERSION	L04162.1	GI:605630			
KEYWORDS					
SOURCE					
ORGANISM	Plasmodium falciparum 3D7				
	Plasmodium falciparum 3D7				
	Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
REFERENCE	1 (bases 1 to 9654)				
AUTHORS	Baheti,S., Alamo,P., Luo,C., Hanra,S., Akawa,M., Carter,R. and Elliott,J.F.				
TITLE	Gene Cloning of a large Plasmodium falciparum sexual stage surface antigen				
JOURNAL	Unpublished (1993)				
COMMENT	Original source text: Plasmodium falciparum (strain 3D7) sexual stage and gametocyte possibly in asexual stages CDNA to mRNA.				
FEATURES					
source	Location/Qualifiers				
	1..9654				
	/organism="Plasmodium falciparum 3D7"				
	/mol_type="mRNA"				
	/strain="3D7"				
	/db_xref="taxon:36329"				
	/map="chromosome 2"				
	/clone="81"				
	/dev_stage="sexual stage and gametocyte possibly in asexual stages"				
	/germline				
	1..9654				
	/gene="s230"				
	1..9654				
	/gene="s230"				
	1..179				
	/gene="s230"				
	179..9586				
	/gene="s230"				
	/codon_start=1				
	/protein_id="AA57559.1"				

```

/translation="MKKIIITLKNLFLIILVYIPEKKKICRVNIKGNIIKDDPKRFB
LFYVSHNLKPTETKKNKCEFTYKNGKIVYLSKRIKRNKDTSVIKTQRTCFHKEG
SSPEWGSKNITCEPIYVGRKERTLDITIIIKKVVNDHVASVSMHNSVNEKNNILIRN
IKENKNDIIONVEKIQDRTYENKDYSDTLLEMPDNDTVNRENFLLTLKRLKMLIF
SSPKRTKTVOKKHKNSPINSLSKAIYMLITPSDSFNDVNRNRNIDDEEMSRDNDV
IDDEEBEEDDEEBEEDDEEBEEDDEEBEEDDEEBEEDDEEBEEDDEEBEEDDEEBE
ESPDNDSDSVARQDMMIRNDEYVEDDDQDVTYDSTIKNEDDEVEGEVEGEVEE
VEGEVEEVEGEVEGEVEGEVEGEVEGEVEGEVEGEVEGEVEGEVEGEVEGEVEGE
OEIYPPGDEEKEDEGESEFTYKSEVDKTLKFIEGGDDVYVDSKVLDDDT
ISRVSKHRTARDEGEYGEVAVEDGENVKIIRSVLSQALPESVADVDLIDLSYET
TESGDVASEDSYDKYASNNTKREYVCDTQDLKPTESGKVCVKEVPEIYKVI
ICPLKGSVEKLYDNIEYVPKSPYVLTKEETKLKEKLSKLYGLLISPTVNEKNN
FKRGVEFTLPVYVNAKATVYFICDNSKEDDKKNGRIGVVEYVYVPGYKNGCALP
DEDEBEKIQNOIEEDBHNKIKMKTFFTONIYKKNIIYCTYKILSGDIGLIPKXN
IKSTCFEEMIYPNKEIKMKNKNSLGNLVNNSVVKNEKKNYFVNOVYHITPSYK
TNLPCSIILKEEBSNLISTYLVVYSINELNFSJDEFESVPIIKKTIQVAKVNV
NKHEDYTCDFDTLDKTPSTANGKLFICRKHKEFDFTLTCNNAKTQPIEILFP
KTLKDKREVLKLDLIOYOMSEKFEKFNTONAKYLNYPYLLIFPNHIGKELKNP
TYKHHKDVYFEQSSVLSPLSADSILKLNFLDQETVCTEKIRYLNLSINELSD
NNTFSYTFQVPPYIDIKEPYPMFGCCKNNKGEINIGVELLISKOBEKIKGNPHSK
LDYFENISDTHECTLHAYENDIIGFNCLETHPNEVEVEDEAEYVDEPENFNV
YKGLNSVDITTLKMAOTVNIINNKTPFTLKIIPPYVLEDEVEISCCITIKQVAKIKV
IITKNDVTDLKREVOSESTLDDKIKYKCHENFNPKNKTFDENVEYTCNIIKENFN
YIOIFCPKADLGIYKNIOMYDIYKPRVPOFKFNNELHKLIPNSEMLHKTKEMLI
LNNEKRVLDLHFVPLPIYIKDIYEENIVCDNKTMMKQGLGKVIYHITSKREOV
KCGSRDNEHMHMFSSYNKTYKNCIIIDAKPKDILGPCPSGTLTKTNCFOAIVHTUN
NNGIILYKNNLANFTYKQPNMEIPALMNDNISRCITVDLKKKYNKYSPLGRV
LRALYKLNIRKFNVTGTQNKYLMYMLHLHSHKNYLKELPHDGRKRPADTDN
PESIIESLSINESNEGPPPTGDVDAHLILEYDWNESYDEOLEVINDISETEL
KDIQYVLYQVNLKAPKLMSAQIHHNNHVCDFSKNNLIVESLKEKEELGNNVNIHC
YALLKPLDLYVKCPTSKDNYEAKVNIISNDNEVELQVLSLEKRPHENETLESKP
GNGDVYVHNGVDTQPLDNPSTFEKYKNIKIDPKREVINYEDTDESEKLESL
PGAIYSPKVLKAKKPPSTYAAVFPVPPYKDIHFKECCKNTEYKEDNOYISGNGIL
HDIISNKRINICDSTNNSILTSSVKNVNETKXCEININNEVFGIICENETNL
DPRKCHIEIYSKNTKTVKREVIYPAIDISLHNSNKKVAAKVPDLYDNKLFSCS
CKTSHNTICTMKVTLNKDEKEEDFKTAQGIKNNVHLNCFNDPELTDNDKIVLC
KIDAEIPESEYIOLPIGTKNVEGVOENEYKFSKPSLVFDNNNDIYVIGKEKE
VJSLSLKGVYGNRIPTPKDNGKKGESIFPIKODTLPKLIINFTIONSINKRG
LIITFRKANSSENSFKCDTGTSTISMEANSQVKEKCTVKIKKGIIFELCKPKGA
IIFPOACSNNVLLSYKSDYEDSEHINYYIHKKDKVYKAPDOVIELMDENRELONTQ
YTGISNTDVLHFKFNGLNPLNFKNYSTAAYAKVDNTSIIINFSCNREPNKHVY
GTWQVEDNRPNIKNENVIKNFLPNIEKVALLDDEEROKIKOOOEEQOQEI
LKQOODELSHDDNKNKHTYLYVDSNEHICDYENESLISLTPDRTKIKOISCKIA
KALDVYTIKCPHTKNFPRKDPYPMNSSLITNDKIVITFDKKNFTYIDPRTKFSK
IYIOSRYGVSULDHNOIKKIHEDNDVHLVYPRHNVLNHNLYNLSALEGVIF
MKSRTGDETRATKNTLLPYDGVSSILIPYKEDITFHLFCGSLTKKKNKNTSLA
LIHIIHSSNNIIGHCDFLYENQNDIAISNNNNNSYIFTHNNKNNLICIISLIP
KTVYGIKCPMKKLNPOTCFDEYVYVVOEDVPSKTIYADKYNSTSKXIGIKLAIISI
NNDEKONTYTYLILPEKFEELIDTKVLACTCDNKYIIMKIKESTMDKIKIDEK
TIKDKICKYVTVTKVATCEIIDIIDSVLKELHVNHYISILSRMDKIIIVYNEKTH
FENFVNPMLKDKVLYNKNPINIEHILPGALITTDIYDRTKIKOYLILIPYVND
IHSLEPNSLSLTKQONITIGYVAKIIFHINOYKEIHGCDPTKYSHLFTSKCP
LRPDDDICVNTIGNTSGFACLSHFELKPNKNCSSVYDNEAKVTKLDELSTKVEL
DHIKONTSGYTLIIYIPNKESTKLKFSCTSSNYSNTIRITFDPNYIPEPOSRAII
KYVLDONKNAFKYLRKL"
9587. 9654
/gene="s230"
3'UTR

```

BASE COUNT 4204 a 976 c 1432 g 3042 t
ORIGIN

Alignment Scores:
Pred. No.: 3.27 Length: 9654
Score: 158.00 Matches: 35
Percent Similarity: 46.02% Conservative: 17
Best Local Similarity: 30.97% Mismatches: 31
Query Match: 15.44% Indels: 30
DB: 3 Gaps: 3

US-10-081-935-4 (1-194) x PFAS230A (1-9654)

OY 57 GYGINlethrProCYsValGluGluAcySerProLeuAspAlaArgIISerValSer 76
DB 1545 GGATATATTTCACCTTC-----CTTCTTTTTC 1516

```

OY 77 AenilleValAlaSpGInCySerlyAlaGlyValProIleAspIleProValAsp 96
DB 1515 TCTACATAT-----TCACCTCTCTTCACTCACTTCACTTCTTCTTCA 1468
OY 97 ThrThAlaAlaProGluProSerGluThrAlaGluProThrAlaGluProGlu 116
DB 1467 CCTACCTTCACTCACTCACTTCACTCACTTCACTTCACTTCACTTCACTTCA 1408
OY 117 ProThAlaGluProThrAlaGluProThrAlaGluProThrAlaGluProGlu 136
DB 1407 CCTACCTTCACTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 1348
OY 137 ProThAlaValProThrGlyThrGlyGlyValProThrGlyThrGlySerPheThr 156
DB 1347 CCTACCTTCACTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 1324
OY 157 ValThrGlyArgProThrAlaSerThrProAlaGluPhe 169
DB 1323 -----CCTACCTTCTCATCTCACTTCACTT 1297

```

RESULT 40
AE001393/c 15148 bp DNA linear INV 04-OCT-2002
LOCUS Plasmidium falciparum 3D7 chromosome 2 section 30 of 73 of the
DEFINITION complete sequence.
ACCESSION AE001393 AE001362
VERSION AE001393.1 GI:3845175
KEYWORDS
SOURCE Plasmidium falciparum 3D7
ORGANISM Plasmidium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 15148)
AUTHORS Gardner,M.J., Tettein,H., Carucci,D.J., Cummings,L.M., Aravind,L., Koonin,E.V., Shalloom,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Jing,D., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.
TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Science 282 (5391), 1126-1132 (1998)
MEDLINE 99021743
PUBMED 9804551

REFERENCE 2 (bases 1 to 15148)
AUTHORS Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W., Carlton,J.M., Paine,A., Nelson,K.E., Bowman,S., Paulsen,I.T., James,S., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A., Kyes,S., Chan,M.-S., Nene,V., Shalloom,S.J., Sub,B., Peterson,J., Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D., Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H., Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I., Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C., Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and Barrett,B.

TITLE Genome sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Nature 419, 498-511 (2002)
REFERENCE 3 (bases 1 to 15148)
AUTHORS Gardner,M.J.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES location/Qualifiers

FEATURES

1..15148 "Plasmidium falciparum 3D7"
/organism="Plasmidium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="2"
<1..44
repeat_region
/rpt_type=tandem
/rpt_unit=(A)n
repeat_region
/rpt_type=tandem

repeat_region 10502..10574
 /rpt_type=tandem
 /rpt_unit=(TAAA)n
 repeat_region 10502..10546

Alignment Scores:

Pred. No.:	4.96	Length:	15148
Score:	158.00	Matches:	35
Percent Similarity:	46.02%	Conservative:	17
Best Local Similarity:	30.97%	Mismatches:	31
Query Match:	15.44%	Indels:	30
DB:	3	Gaps:	3

US-10-081-935-4 (1-194) x AE001393 (1-15148)

```

QY      57  GlyGlnIleThrProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSer 76
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
          2236 GGATATATTTACCTTC-----CTTCTTTTCA 2207

QY      77  AsnIleValValAspGlnCysSerLysAlaGlyValProIleAspIleProProValAsp 96
          ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|  ::|
          2206 TCTACATAT-----TCACCTCTCTCTTCACTTCACTTCACTTCTTCA 2159

QY      97  ThrThraIaIaProGluProSerGluThraIaGluProThraIaGluProThrgluGlu 116
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
          2158 CCACTCTTCACCTACCTTCACTCACTCACTCTTCTTCACTTCACTTCACTTCA 2099

QY      117 ProThraIaGluProThraIaGluProThraIaGluProThraIaGluProThrgluGlu 136
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
          2098 CCACTCTTCACCTACCTTCACTCACTCACTCTTCACTTCACTTCACTTCACTTCA 2039

QY      137 ProThraIaValProThrglyThrglyGlyGlyValProThrglyThrglySerPheThr 156
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
          2038 CCACTCTTCACCTACCTCTTCA----- 2015

QY      157 ValThrglyArgProThraIaSerThrProAlaGluPhe 169
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
          2014 -----CCTACCTCTTCATCTACATCTTCATTT 1988

```

Search completed: December 6, 2003, 03:09:16
 Job time : 2869 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 13, 2003, 11:51:07 / Search time 52 Seconds

(without alignments)
152.786 Million cell updates/sec

Title: US-10-081-935-2

Perfect score: 82

Sequence: 1 M0FSHALITALVAAGLASA 18

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 segs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPRO_pool/US10081935/runat_12112003_144407_3656/app_query.fasta_1.199
-DB=Issued_Patents_NA -QFMT=faststep -SUFFIX=rm1 -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10081935 @CGN 1.1 44 @runat_12112003_144407_3656 -NCPU=6 -ICPU=3
-NO MAP -LARGEBUFFER -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOC
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	57.3	4403765	3 US-09-103-840A-2	Sequence 2, Appli
2	47	57.3	4411529	3 US-09-103-840A-1	Sequence 1, Appli
3	44	53.7	248	1 US-08-525-697-3	Sequence 3, Appli
4	44	53.7	1302	1 US-08-525-697-1	Sequence 1, Appli
5	44	53.7	10091	3 US-09-058-489-34	Sequence 34, Appli
6	42	51.2	471	4 US-09-252-991A-8612	Sequence 8612, Ap
7	42	51.2	504	4 US-09-134-001C-959	Sequence 959, App
8	42	51.2	1781	2 US-08-389-564B-21	Sequence 21, Appl
9	42	51.2	1781	2 US-08-389-564B-21	Sequence 22, Appl
10	42	51.2	1781	3 US-08-466-047B-22	Sequence 21, Appl
11	42	51.2	1781	3 US-08-466-047B-22	Sequence 22, Appl
12	42	51.2	2218	2 US-08-389-564B-18	Sequence 18, Appl

C 13	42	51.2	2218	3	US-09-066-597-1	Sequence 1, Appli
C 14	42	51.2	2218	3	US-08-466-047B-18	Sequence 18, Appl
C 15	42	51.2	2266	2	US-08-389-564B-20	Sequence 20, Appl
C 16	42	51.2	2266	3	US-08-466-047B-20	Sequence 20, Appl
C 17	42	51.2	4176	3	US-09-254-733-1	Sequence 1, Appli
C 18	42	51.2	8075	3	US-08-374-077C-1	Sequence 1, Appli
C 19	42	51.2	8075	3	US-08-895-590-1	Sequence 1, Appli
C 20	42	51.2	8075	4	US-09-539-879A-1	Sequence 1, Appli
C 21	42	51.2	8096	3	US-09-058-489-33	Sequence 33, Appl
C 22	41	50.0	528	4	US-09-252-991A-9137	Sequence 9137, Ap
C 23	41	50.0	972	4	US-09-252-991A-15702	Sequence 15702, A
C 24	41	50.0	975	4	US-09-252-991A-15674	Sequence 15674, A
C 25	41	50.0	1056	4	US-09-252-991A-8728	Sequence 8728, Ap
C 26	41	50.0	1218	4	US-09-552-991A-322-3	Sequence 3, Appli
C 27	41	50.0	1230	4	US-09-328-352-1918	Sequence 1918, Ap
C 28	41	50.0	1865	4	US-09-252-991A-15656	Sequence 15656, A
C 29	41	50.0	1566	4	US-09-252-991A-9220	Sequence 9220, Ap
C 30	41	50.0	1710	4	US-09-252-991A-9057	Sequence 9057, Ap
C 31	41	50.0	2123	4	US-09-724-864-18	Sequence 18, Appl
C 32	41	50.0	5045	3	US-09-390-721-1	Sequence 1, Appli
C 33	41	50.0	5045	3	US-09-390-721-3	Sequence 3, Appli
C 34	41	50.0	36412	4	US-08-311-731A-132	Sequence 132, App
C 35	40	48.8	304	4	US-09-313-294A-2891	Sequence 2891, Ap
C 36	40	48.8	465	4	US-09-252-991A-16275	Sequence 16275, A
C 37	40	48.8	504	4	US-09-512-342-19	Sequence 19, Appl
C 38	40	48.8	669	4	US-09-252-991A-3703	Sequence 3703, Ap
C 39	40	48.8	717	3	US-08-913-014A-7	Sequence 7, Appli
C 40	40	48.8	765	4	US-09-252-991A-9931	Sequence 9931, Ap
C 41	40	48.8	795	3	US-08-913-014A-8	Sequence 8, Appli
C 42	40	48.8	951	4	US-09-252-991A-16386	Sequence 16386, A
C 43	40	48.8	1053	4	US-09-252-991A-6421	Sequence 6421, Ap
C 44	40	48.8	1092	4	US-09-252-991A-16212	Sequence 16212, A
C 45	40	48.8	1179	4	US-09-252-991A-10081	Sequence 10081, A

ALIGNMENTS

RESULT 1
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
Alignment Scores:
Pred. No.: 1.85e+05
Score: 47.00
Percent Similarity: 92.86%
Best Local Similarity: 64.29%
Query Match: 57.32%
DB: 3
Length: 4403765
Matches: 9
Conservative: 4
Mismatches: 1
Indels: 0
Gaps: 0
US-10-081-935-2 (1-18) x US-09-103-840A-2 (1-4403765)

OY 3 PheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAla 16
DB 1218600 TTGCGCACGCCCTGGTGTGATGCTGTCGCCGCTGAATTGGCG 1218641

RESULT 2
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-2007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 1.85e+05 Length: 4411529
Score: 47.00 Matches: 9
Percent Similarity: 92.86% Conservative: 4
Best Local Similarity: 64.29% Mismatches: 1
Query Match: 57.32% Indels: 0
DB: 3 Gaps: 0

US-10-081-935-2 (1-18) x US-09-103-840A-1 (1-4411529)

OY 3 PheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAla 16
DB 1219072 TTGCGCACGCCCTGGTGTGATGCTGTCGCCGCTGAATTGGCG 1219113

RESULT 3
US-08-525-697-3
Sequence 3, Application US/08525697
Patent No. 5795764
GENERAL INFORMATION:
APPLICANT: Christgau, Stephan
APPLICANT: Andersen, Lene N
APPLICANT: Kauppinen, Sakari
APPLICANT: Heidt-Hansen, Hans P
APPLICANT: Dalboege, Henrik
TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 57957640 No. 5795764disk of No. 5795764th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,697
FILING DATE: 21-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4004.204-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 0486/93
FILING DATE: 30-APR-1993
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 248 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-525-697-3

Alignment Scores:
Pred. No.: 3.04 Length: 248
Score: 44.00 Matches: 8
Percent Similarity: 77.78% Conservative: 6
Best Local Similarity: 44.44% Mismatches: 4
Query Match: 53.66% Indels: 0
DB: 1 Gaps: 0

US-10-081-935-2 (1-18) x US-08-525-697-3 (1-248)

OY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAla 18
DB 24 ATGACCTTCTTCACATGCTCTCCACCTCGCCACCGCTGGGGGTGCGACGCGCT 77

RESULT 4
US-08-525-697-1
Sequence 1, Application US/08525697
Patent No. 5795764
GENERAL INFORMATION:
APPLICANT: Christgau, Stephan
APPLICANT: Andersen, Lene N
APPLICANT: Kauppinen, Sakari
APPLICANT: Heidt-Hansen, Hans P
APPLICANT: Dalboege, Henrik
TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 57957640 No. 5795764disk of No. 5795764th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,697
FILING DATE: 21-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4004.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 0486/93
FILING DATE: 30-APR-1993
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1302 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

```

; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 61..1192
US-08-525-697-1

Alignment Scores:
Pred. No.: 25
Score: 44.00
Percent Similarity: 77.78%
Best Local Similarity: 44.44%
Query Match: 53.66%
DB: 1
Gaps: 0

US-10-081-935-2 (1-18) x US-08-525-697-1 (1-1302)

QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 61 ATGAAGCTTCTTCATGCTCTCCACGCTCGCCAGCCTGGGGGTGGCGAGCGCT 114

RESULT 5
US-09-058-489-34
; Sequence 34, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WH197-08PA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; PRIOR FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 10091
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-34

Alignment Scores:
Pred. No.: 339
Score: 44.00
Percent Similarity: 76.47%
Best Local Similarity: 52.94%
Query Match: 53.66%
DB: 3
Gaps: 0

US-10-081-935-2 (1-18) x US-09-058-489-34 (1-10091)

QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSer 17
Db 3483 CTTCAACAAATCATGCTTGTACTTGTGTAGCAGAAACCTTGCAACC 3533

RESULT 6
US-09-252-991A-8612
; Sequence 8612, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
```

```

; SEQ ID NO 8612
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8612

Alignment Scores:
Pred. No.: 16.8
Score: 42.00
Percent Similarity: 76.47%
Best Local Similarity: 41.18%
Query Match: 51.22%
DB: 4
Gaps: 0

US-10-081-935-2 (1-18) x US-09-252-991A-8612 (1-471)

QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSer 17
Db 70 ATGAATTCACGATTCATCGCGGCATCGTTGGCCGCCACTGTCACAC 120

RESULT 7
US-09-134-001C-959/C
; Sequence 959, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucet-Re-Stein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 959
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-959

Alignment Scores:
Pred. No.: 18.3
Score: 42.00
Percent Similarity: 82.35%
Best Local Similarity: 58.82%
Query Match: 51.22%
DB: 4
Gaps: 1

US-10-081-935-2 (1-18) x US-09-134-001C-959 (1-504)

QY 2 GlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
Db 134 CAATTTCACATCTCTTCTACGCTTAATCGCA-----TTGGCCCTCATCT 90

RESULT 8
US-08-389-564B-21/C
; Sequence 21, Application US/08389564B
; Patent No. 5989870
; GENERAL INFORMATION:
; APPLICANT: Nakari, Tiina H.
; APPLICANT: Onnela, Maija-Leena
; APPLICANT: Iim n, Marja H.
; APPLICANT: Penttil, Merja E.
; TITLE OF INVENTION: A METHOD FOR CLONING ACTIVE PROMOTERS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scerne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```

;
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,564B
; FILING DATE: 16-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/932,485
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/496,155
; FILING DATE: 19-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/044,077
; FILING DATE: 29-APR-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 86 10600
; FILING DATE: 30-APR-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, GRANT E.
; REGISTRATION NUMBER: 41,264
; REFERENCE/DOCKET NUMBER: 1716.008000G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1781 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
;
; US-08-389-564B-21
;
; Alignment Scores:
; Pred. No.: 91.3 length: 1781
; Score: 42.00 Matches: 7
; Percent Similarity: 84.62% Conservative: 4
; Best Local Similarity: 53.85% Mismatches: 2
; Query Match: 51.22% Indels: 0
; DB: 2 Gaps: 0
;
; US-10-081-935-2 (1-18) x US-08-389-564B-21 (1-1781)
;
; QY 3 pHeSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeu 15
;
; Db 1031 TTAGCCACTGTGATGTACATTATTAGCCGCTGTATA 993
;
; RESULT 9
; US-08-389-564B-22/C
; Sequence 22, Application US/08389564B
; Patent No. 5989870
; GENERAL INFORMATION:
; APPLICANT: Nakari, Tiina H.
; APPLICANT: Onnela, Maija-Leena
; APPLICANT: Ilm n, Marja H.
; APPLICANT: Penttil, Merja E.
; TITLE OF INVENTION: A METHOD FOR CLONING ACTIVE PROMOTERS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; SOFTWARE: Patentin Release #1.0, Version #1.25
; COMPUTER: IBM PC compatible
```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,564B
; FILING DATE: 16-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/932,485
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/496,155
; FILING DATE: 19-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/044,077
; FILING DATE: 29-APR-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 86 10600
; FILING DATE: 30-APR-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, GRANT E.
; REGISTRATION NUMBER: 41,264
; REFERENCE/DOCKET NUMBER: 1716.008000G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1781 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
;
; US-08-389-564B-22
;
; Alignment Scores:
; Pred. No.: 91.3 length: 1781
; Score: 42.00 Matches: 7
; Percent Similarity: 84.62% Conservative: 4
; Best Local Similarity: 53.85% Mismatches: 2
; Query Match: 51.22% Indels: 0
; DB: 2 Gaps: 0
;
; US-10-081-935-2 (1-18) x US-08-389-564B-22 (1-1781)
;
; QY 3 pHeSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeu 15
;
; Db 1031 TTAGCCACTGTGATGTACATTATTAGCCGCTGTATA 993
;
; RESULT 10
; US-08-466-047B-21/C
; Sequence 21, Application US/08466047B
; Patent No. 6011147
; GENERAL INFORMATION:
; APPLICANT: Nakari, Tiina H.
; APPLICANT: Onnela, Maija-Leena
; APPLICANT: Ilm n, Marja H.
; APPLICANT: Nevalainen, Kaisa Milja Helena
; APPLICANT: Penttil, Merja E.
; TITLE OF INVENTION: Fungal Promoters Active In The Presence
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,047B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/389,564
FILING DATE: 16-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/932,564
FILING DATE: 19-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/496,155
FILING DATE: 19-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/044,077
FILING DATE: 29-APR-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 86 10600
FILING DATE: 30-APR-1986
ATTORNEY/AGENT INFORMATION:
NAME: REED, GRANT E.
REGISTRATION NUMBER: 41,264
REFERENCE/DOCKET NUMBER: 1716.008000H
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1781 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
US-08-466-047B-21

Alignment Scores:
Pred. No.: 91.3 Length: 1781
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
Gaps: 0
DB: 3

US-10-081-935-2 (1-18) x US-08-466-047B-21 (1-1781)
QY 3 PheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeu 15
Db 1031 TTAGCCACTGTGATTGTACATTATTAGCCGCTGTATA 993

RESULT 11
US-08-466-047B-22/C
Sequence 22, Application US/08466047B
Patent No. 6011147
GENERAL INFORMATION:
APPLICANT: Nakari, Tina H.
APPLICANT: Onnela, Maija-Leena
APPLICANT: Ilm n, Marja H.
APPLICANT: Nevalainen, Kaisa
APPLICANT: Penttil, Merja E.
TITLE OF INVENTION: Fungal Promoters Active In The Presence
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,047B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/389,564
FILING DATE: 16-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/932,564
FILING DATE: 19-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/496,155
FILING DATE: 19-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/044,077
FILING DATE: 29-APR-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 86 10600
FILING DATE: 30-APR-1986
ATTORNEY/AGENT INFORMATION:
NAME: REED, GRANT E.
REGISTRATION NUMBER: 41,264
REFERENCE/DOCKET NUMBER: 1716.008000H
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2540
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 1781 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
US-08-466-047B-22

Alignment Scores:
Pred. No.: 91.3 Length: 1781
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
Gaps: 0
DB: 3

US-10-081-935-2 (1-18) x US-08-466-047B-22 (1-1781)
QY 3 PheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeu 15
Db 1031 TTAGCCACTGTGATTGTACATTATTAGCCGCTGTATA 993

RESULT 12
US-08-389-564B-18/C
Sequence 18, Application US/08389564B
Patent No. 5989870
GENERAL INFORMATION:
APPLICANT: Nakari, Tina H.
APPLICANT: Onnela, Maija-Leena
APPLICANT: Ilm n, Marja H.
APPLICANT: Penttil, Merja E.
TITLE OF INVENTION: A METHOD FOR CLONING ACTIVE PROMOTERS
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/389,564B
: FILING DATE: 16-FEB-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/932,485
: FILING DATE: 19-AUG-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/496,155
: FILING DATE: 19-MAR-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/044,077
: FILING DATE: 29-APR-1987
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 86 10600
: FILING DATE: 30-APR-1986
: ATTORNEY/AGENT INFORMATION:
: NAME: REED, GRANT E.
: REGISTRATION NUMBER: 41,264
: REFERENCE/DOCKET NUMBER: 1716.008000G
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: both
: MOLECULE TYPE: CDNA
:
: US-08-389-564B-18
:
: Alignment Scores:
: Pred. No.: 121 Length: 2218
: Score: 42.00 Matches: 7
: Percent Similarity: 84.62% Conservative: 4
: Best Local Similarity: 53.85% Mismatches: 2
: Query Match: 51.22% Indels: 0
: DB: 2 Gaps: 0
:
: US-10-081-935-2 (1-18) x US-08-389-564B-18 (1-2218)
:
: QY 3 PheseRHisAlaLeuIleAlaLeuValAlaAcyLeu 15
: Db 1466 TTAGCCACTTGATTGTACATTTATTAGCCGCTGTATA 1428
:
: RESULT 13
: US-09-066-597-1/c
: Sequence 1, Application US/09066597
: Patent No. 6001595
: GENERAL INFORMATION:
: APPLICANT: IIm n, Marija
: APPLICANT: Onnela, Marija-Leena
: APPLICANT: Penttil, Merja
: TITLE OF INVENTION: New Promoters and Uses Thereof
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, NW, Suite 600
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/066,597
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
```

```

: APPLICATION NUMBER: US 60/032,156
: FILING DATE: 29-NOV-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/032,959
: FILING DATE: 13-DEC-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/040,140
: FILING DATE: 10-MAR-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/980,061
: FILING DATE: 26-NOV-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FI97/00742
: FILING DATE: 01-DEC-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Kim, Judith U.
: REGISTRATION NUMBER: 40,679
: REFERENCE/DOCKET NUMBER: 1716.0660004/MAC/JUK
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
:
: US-09-066-597-1
:
: Alignment Scores:
: Pred. No.: 121 Length: 2218
: Score: 42.00 Matches: 7
: Percent Similarity: 84.62% Conservative: 4
: Best Local Similarity: 53.85% Mismatches: 2
: Query Match: 51.22% Indels: 0
: DB: 3 Gaps: 0
:
: US-10-081-935-2 (1-18) x US-09-066-597-1 (1-2218)
:
: QY 3 PheseRHisAlaLeuIleAlaLeuValAlaAcyLeu 15
: Db 1466 TTAGCCACTTGATTGTACATTTATTAGCCGCTGTATA 1428
:
: RESULT 14
: US-08-466-047B-18/c
: Sequence 18, Application US/08466047B
: Patent No. 6011147
: GENERAL INFORMATION:
: APPLICANT: Nakari, Tiina H.
: APPLICANT: Onnela, Marija-Leena
: APPLICANT: IIm n, Marija H.
: APPLICANT: Nevalainen, Kaisu Marija Helena
: APPLICANT: Penttil, Merja E.
: TITLE OF INVENTION: Fungal Promoters Active In The Presence
: OF Glucose
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/466,047B
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,564
; FILING DATE: 16-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/932,564
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/496,155
; FILING DATE: 19-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/044,077
; FILING DATE: 29-APR-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 86 10600
; FILING DATE: 30-APR-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, GRANT E.
; REGISTRATION NUMBER: 41,264
; REFERENCE/DOCKET NUMBER: 1716.008000H
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; US-08-466-047B-18

Alignment Scores:
Pred. No.: 121 Length: 2218
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
Gaps: 0

US-10-081-935-2 (1-18) x US-08-466-047B-18 (1-2218)

QY 3 PheSerHisAlaLeuIleAlaLeuValAlaIleGlyLeu 15
Db 1466 TTTAGCCACTTGATGTGCAATTATTAGCCGCTGGTATA 1428

RESULT 15
US-08-389-564B-20/C
; Sequence 20, Application US/08389564B
; Patent No. 5989870
; GENERAL INFORMATION:
; APPLICANT: Nakari, Tiina H.
; APPLICANT: Onnela, Maija-Leena
; APPLICANT: Ilm n, Marja H.
; APPLICANT: Penttil, Merja E.
; TITLE OF INVENTION: A METHOD FOR CLONING ACTIVE PROMOTERS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,564B
; FILING DATE: 16-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: US 07/932,485
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/496,155
; FILING DATE: 19-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/044,077
; FILING DATE: 29-APR-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 86 10600
; FILING DATE: 30-APR-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, GRANT E.
; REGISTRATION NUMBER: 41,264
; REFERENCE/DOCKET NUMBER: 1716.008000G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; US-08-389-564B-20

Alignment Scores:
Pred. No.: 124 Length: 2266
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
Gaps: 0

US-10-081-935-2 (1-18) x US-08-389-564B-20 (1-2266)

QY 3 PheSerHisAlaLeuIleAlaLeuValAlaIleGlyLeu 15
Db 1516 TTTAGCCACTTGATGTGCAATTATTAGCCGCTGGTATA 1478

RESULT 16
US-08-466-047B-20/C
; Sequence 20, Application US/08466047B
; Patent No. 6011417
; GENERAL INFORMATION:
; APPLICANT: Nakari, Tiina H.
; APPLICANT: Onnela, Maija-Leena
; APPLICANT: Ilm n, Marja H.
; APPLICANT: Nevalainen, Kaisa Milja Helena
; APPLICANT: Penttil, Merja E.
; TITLE OF INVENTION: Fungal Promoters Active In The Presence
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,047B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,564
; FILING DATE: 16-FEB-1995
```



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/932,564
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/496,155
; FILING DATE: 19-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/044,077
; FILING DATE: 29-APR-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 86 10600
; FILING DATE: 30-APR-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, GRANT E.
; REGISTRATION NUMBER: 41,264
; REFERENCE/DOCKET NUMBER: 1716.008000H
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; US-08-466-047B-20

Alignment Scores:
Pred. No.: 124 Length: 2266
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
DB: 3 Gaps: 0

US-10-081-935-2 (1-18) x US-08-466-047B-20 (1-2266)
QY 3 PhseRHsAlaleuileAlaleuValAlaAlaGlyLeu 15
Db 1516 TTTAGCCACTTGTGTGTACATTATTAGCGCGTGTATA 1478

RESULT 17
US-09-254-733-1/c
; Sequence 1, Application US/09254733
; Patent No. 6277596
; GENERAL INFORMATION:
; APPLICANT: WATANABE, MANABU
; APPLICANT: MORIYA, TATSUKI
; APPLICANT: AOYAGI, KAORU
; APPLICANT: SUMIDA, NAOMI
; APPLICANT: MURAKAMI, TAKESHI
; TITLE OF INVENTION: REGULATORY SEQUENCE OF CELLULASE cbh1 GENES ORIGINATING
; TITLE OF INVENTION: IN TRICHODERMA VIRIDE AND SYSTEM FOR MASS-PRODUCING
; FILE REFERENCE: 99-0266*/LC(IMC)/00144
; CURRENT APPLICATION NUMBER: US/09/254,733
; FILING DATE: 1999-05-07
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4176
; TYPE: DNA
; ORGANISM: TRICHODERMA VIRIDE MC300-1
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (1438) .. (1488)
; NAME/KEY: CDS
; LOCATION: (1438) .. (1488)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (1489) .. (3108)
```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1489) .. (1898)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (1899) .. (1965)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1966) .. (2662)
; FEATURE:
; NAME/KEY: intron
; LOCATION: (2663) .. (2724)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2725) .. (3108)
; US-09-254-733-1

Alignment Scores:
Pred. No.: 270 Length: 4176
Score: 42.00 Matches: 7
Percent Similarity: 84.62% Conservative: 4
Best Local Similarity: 53.85% Mismatches: 2
Query Match: 51.22% Indels: 0
DB: 3 Gaps: 0

US-10-081-935-2 (1-18) x US-09-254-733-1 (1-4176)
QY 3 PhseRHsAlaleuileAlaleuValAlaAlaGlyLeu 15
Db 708 TTTAGCCACTTATTGTGTACATTATTAGCGCGTGTATA 670

RESULT 18
US-08-374-077C-1/c
; Sequence 1, Application US/08374077C
; Patent No. 6027912
; GENERAL INFORMATION:
; APPLICANT: Hall, Linda M.
; APPLICANT: Ren, Dejian
; APPLICANT: Zheng, Wei
; APPLICANT: Dubald, Manuel Marcel Paul
; TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha1
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
; STREET: 699 Prince Street
; City: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,077C
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm M.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 022650-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8075 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

FEATURE:
NAME/KEY: CDS
LOCATION: 157..7704
US-08-374-077C-1

Alignment Scores:
Pred. No.: 624
Score: 42.00
Percent Similarity: 83.33%
Best Local Similarity: 44.44%
Query Match: 51.22%
DB: 3
Indels: 0
Gaps: 0

US-10-081-935-2 (1-18) x US-08-374-077C-1 (1-8075)

OY 1 MetGlnPheSerHisAlaLeuValAlaIaGlyLeuAlaSerAla 18
Db 1917 TTGGCGTTTCCAGCGCGTTGTCTTTGGCCAGCGCACCTGCACGACG 1864

RESULT 19
US-08-895-590-1/C
Sequence 1, Application US/08895590
Patent No. 6207410
GENERAL INFORMATION:
APPLICANT: Hall, Linda M.
APPLICANT: Ren, Dejian
APPLICANT: Zheng, Wei
APPLICANT: Dubald, Manuel Marcel Paul
TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,590
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/374,888
FILING DATE: 19-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm M.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 022650-263
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8075 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 157..7704
US-08-895-590-1

Alignment Scores:
Pred. No.: 624
Score: 42.00
Percent Similarity: 83.33%
Best Local Similarity: 44.44%
Query Match: 51.22%
DB: 3
Indels: 0
Gaps: 0

Query Match: 51.22%
DB: 3
Indels: 0
Gaps: 0

US-10-081-935-2 (1-18) x US-08-895-590-1 (1-8075)

OY 1 MetGlnPheSerHisAlaLeuValAlaIaGlyLeuAlaSerAla 18
Db 1917 TTGGCGTTTCCAGCGCGTTGTCTTTGGCCAGCGCACCTGCACGACG 1864

RESULT 20
US-09-539-879A-1/C
Sequence 1, Application US/09539879A
Patent No. 6436627
GENERAL INFORMATION:
APPLICANT: Hall, Linda M.
APPLICANT: Ren, Dejian
APPLICANT: Zheng, Wei
APPLICANT: Dubald, Manuel Marcel Paul
TITLE OF INVENTION: Genes Encoding an Invertebrate Alpha
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/539,879A
FILING DATE: 31-Mar-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/111,865
FILING DATE: <Unknown>
APPLICATION NUMBER: US 08/374,077
FILING DATE: 19-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm M.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 022650-264
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8075 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 157..7704
US-09-539-879A-1

Alignment Scores:
Pred. No.: 624
Score: 42.00
Percent Similarity: 83.33%
Best Local Similarity: 44.44%
Query Match: 51.22%
DB: 4
Indels: 0
Gaps: 0

US-10-081-935-2 (1-18) x US-09-539-879A-1 (1-8075)

OY 1 MetGlnPheSerHisAlaLeuValAlaIaGlyLeuAlaSerAla 18

Db
1917 TTGCGTTTCCACGGCGTTGTCTTTGGCCAGCGGCACCTGCACAGC 1864

```

RESULT 2)
US-09-056-489-33
Sequence 33, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
APPLICANT: Page, David
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
TITLE OF INVENTION: the Y Chromosome
FILE REFERENCE: WH197-08PA
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 33
LENGTH: 8096
TYPE: DNA
ORGANISM: Human
US-09-058-489-33

```

Alignment Scores:	
Pred. No.:	626
Score:	42.00
Percent Similarity:	70.59%
Best Local Similarity:	47.06%
Query Match:	51.22%
DB:	3
Length:	8036
Matches:	8
Conservative:	4
Mismatches:	5
Indels:	0
Gaps:	0

US-10-081-935-2 (1-18) x US-09-058-489-33 (1-80936)

Qy 1 MetClnphseRHisIslaIeuLleAlaValAlaAlaGlyLeuAlaSer 17
:::|||||
1854 CTTCAACAGCATATGCCCTAGTACTTGTCGACGAGAAACCTGTAACT 1904
:::|||||

```

RESULT 22
US-09-252-991A-9137/c
: Sequence 9137, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 9137
: LENGTH: 528
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9137

```

Alignment Scores:	
Pred. No.:	30.4
Score:	41.00
Percent Similarity:	61.11%
Best Local Similarity:	44.44%
Query Match:	50.00%
DB:	4
Length:	528
Matches:	8
Conservative:	3
Mismatches:	3
Indels:	0
Gaps:	0

US-10-081-935-2 (1-18) x US-09-252-991A-9137 (1-528)
Qy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18

Db 437 CTGCAGTTGACACCACTACTTGGCGGAGTGGCGGCGGCGGCTTGGCGACGCG 384

```

RESULT 23
US-09-252-991A-15702
; Sequence 15702, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 3142
; SEQ ID NO 15702
; LENGTH: 972
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (669)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-15702

```

Alignment Scores:	
Pred. No.:	66.1
Score:	41.00
Percent Similarity:	72.22%
Best Local Similarity:	61.1%
Query Match:	50.00%
DB:	4
Length:	972
Matches:	11
Conservative:	2
Mismatches:	5
Indels:	0
Gaps:	0

Oy 1 MetClnPhseSerHisAlaIeuIleAlaIeuValAlaAlaGlyLeuAlaAsnSerAla 18
| | | | | : | | | | | : | | | | | : | | | | |
Db 151 ATTCAGACGGCCGCCCGCTGTATCCGCCTGTCGGCGCCGGAGACAGCATGGC 204

US-10-081-935-2 (1-18) x US-09-252-991A-15702 (1-972)

```

RESULT 24
US-09-252-991A-15674
: Sequence 15674, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: PRIORITY FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 15674
: LENGTH: 975
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (571)
: OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-15674

```

Alignment Scores:	
Pred. No.:	66.3
Score:	41.00
Percent Similarity:	72.22%
Length:	975
Matches:	11
Conservative:	2

Alignment Scores:

Pred. No.: 113 Length: 1485
Score: 41.00 Matches: 11
Percent Similarity: 72.22% Conservative: 2
Best Local Similarity: 61.11% Mismatches: 5
Query Match: 50.00% Indels: 0
DB: 4 Gaps: 0

US-10-081-935-2 (1-18) x US-09-252-991A-15656 (1-1485)

QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
DB 933 CTGCAGGCGCGCGCGCTGATCGCGCTGCGCGCGGAGAGCATGCGG 880

RESULT 29

US-09-252-991A-9220/C
Sequence 9220, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 9220
LENGTH: 1566
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9220

Alignment Scores:

Pred. No.: 121 Length: 1566
Score: 41.00 Matches: 8
Percent Similarity: 61.11% Conservative: 3
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 50.00% Indels: 0
DB: 4 Gaps: 0

US-10-081-935-2 (1-18) x US-09-252-991A-9220 (1-1566)

QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
DB 979 CTGCAGGTTGAACACCACTGCTGCGGATCGCGCGGCTCGGCAGCGG 926

RESULT 30

US-09-252-991A-9057/C
Sequence 9057, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 9057
LENGTH: 1710
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9057

Alignment Scores:

Pred. No.: 136 Length: 1710
Score: 41.00 Matches: 8
Percent Similarity: 61.11% Conservative: 3
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 50.00% Indels: 0
DB: 4 Gaps: 0

US-10-081-935-2 (1-18) x US-09-252-991A-9057 (1-1710)

QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
DB 1143 CTGCAGGTTGAACACCACTGCTGCGGATCGCGCGGCTCGGCAGCGG 1090

RESULT 31

US-09-724-864-18
Sequence 18, Application US/09724864
Patent No. 6380362
GENERAL INFORMATION:
APPLICANT: Watson, James D.
TITLE OF INVENTION: Polynucleotides, polypeptides expressed
FILE REFERENCE: 11000.1050U1
CURRENT APPLICATION NUMBER: US/09/724,864
CURRENT FILING DATE: 2000-11-28
PRIOR FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 2123
TYPE: DNA
ORGANISM: Mouse
US-09-724-864-18

Alignment Scores:

Pred. No.: 179 Length: 2123
Score: 41.00 Matches: 8
Percent Similarity: 66.67% Conservative: 4
Best Local Similarity: 44.44% Mismatches: 6
Query Match: 50.00% Indels: 0
DB: 4 Gaps: 0

US-10-081-935-2 (1-18) x US-09-724-864-18 (1-2123)

QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAla 18
DB 870 CTGCCTTGAGCCACCACTGCTGCTGCTGCGCAATGCGTGTGTACAGCA 923

RESULT 32

US-09-390-721-1/C
Sequence 1, Application US/09390721
Patent No. 6197591
GENERAL INFORMATION:
APPLICANT: STUTZMAN-ENGWALL, KIM J.
TITLE OF INVENTION: STREPTOMYCES AVERMITILIS REGULATORY GENES FOR INCREASED
FILE REFERENCE: PC9944A
CURRENT APPLICATION NUMBER: US/09/390,721
CURRENT FILING DATE: 1999-09-07
EARLIER APPLICATION NUMBER: 60/100,134
EARLIER FILING DATE: 1998-09-14
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 5045
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1112)..(2317)
OTHER INFORMATION: aver1 ORF

```
US-09-390-721-1
Alignment Scores:
Pred. No.: 537
Score: 41.00
Percent Similarity: 70.59%
Best Local Similarity: 47.06%
Query Match: 50.00%
DB: 3
Gaps: 0

US-10-081-935-2 (1-18) x US-09-390-721-1 (1-5045)
QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaIleGlyLeuAlaSer 17
DB 4382 CTGCAACTGACACACCGCTTGCTGCACTTCGTGAGCGCGGCTGCGCC 4332

RESULT 33
US-09-390-721-3/c
; Sequence 3, Application US/09390721
; Patent No. 6197591
; GENERAL INFORMATION:
; APPLICANT: STUTZMAN-ENGWALL, KIM J.
; APPLICANT: PRICE, BRENDA S.
; TITLE OF INVENTION: STREPTOMYCES AVERMITILIS REGULATORY GENES FOR INCREASED
; FILE REFERENCE: PC9944A
; CURRENT APPLICATION NUMBER: US/09/390,721
; EARLIER FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 5045
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; NAME/KEY: CDS
; FEATURE:
; LOCATION: (2314)..(3021)
; OTHER INFORMATION: aver2 ORF
US-09-390-721-3

Alignment Scores:
Pred. No.: 537
Score: 41.00
Percent Similarity: 70.59%
Best Local Similarity: 47.06%
Query Match: 50.00%
DB: 3
Gaps: 0

US-10-081-935-2 (1-18) x US-09-390-721-3 (1-5045)
QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaIleGlyLeuAlaSer 17
DB 4382 CTGCAACTGACACACCGCTTGCTGCACTTCGTGAGCGCGGCTGCGCC 4332

RESULT 34
US-08-311-731A-132
; Sequence 132, Application US/08311731A
; Patent No. 6583286
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-1
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA

US-10-081-935-2 (1-18) x US-08-311-731A-132 (1-36412)
QY 4 SerHisAlaLeuIleAlaLeuValAlaIleGlyLeuAlaSer 17
DB 8097 ACTCAGCGCGCTTCGCGCACTCATCGTCTGCGCTGAGCTCG 8138

RESULT 35
US-09-313-294A-2891/c
; Sequence 2891, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 2891
; LENGTH: 304
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID NO. 6476212 700553448H1
; NAME/KEY: unsure
; LOCATION: 138, 178
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-2891

Alignment Scores:
Pred. No.: 23.6
Score: 40.00
Length: 304
Matches: 9
```


AUTHORS Gardner, M.J. and Kirkland, T.
TITLE Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL Unpublished
COMMENT Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org.

FEATURES
source Location/Qualifiers
1..737

/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAA67"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 136 a 271 c 168 g 162 t
ORIGIN

Alignment Scores:
Pred. No.: 9.2e-69 Length: 737
Score: 1010.00 Matches: 194
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 0
Query Match: 98.73% Indels: 1
DB: 10 Gaps: 0

US-10-081-935-4 (1-194) x BF251103 (1-737)

OY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
DB 99 ATGCAATTCTCTCACGCTCTCATCGCTCGTGGCGGCGCTCGCAAGTCCAGGCTC 158
OY 21 ProAapIleProProCysAlaLeuAsnCysPheValGluAlaLeuGlyAsnAspGlyCys 40
DB 159 CCAGACATCCCACTTGGCTCTCAACTGCTTGTGAGGCTCTCGCAACGATGGCTGC 218
OY 41 ThrArgLeuThrAspPheLysCysHisCysSerLysProGluLeuProGlyGlnIleThr 60
DB 219 ACTGCTTGACGCACTTCAAGTCCACTGCTCCAAAGCTGAGCTCCAGAGACAGATCACT 278
OY 61 ProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleValVal 80
DB 279 CCTGCGTTGAGAGAGCGCTGCCCTCTCGACGCCGATCTCGCTCCCAACATCGCTT 338
OY 81 AspGlnCysSerLysAlaGlyValProIleAapIleProProValAspThrThrAlaAla 100
DB 339 GACCAAGTCTCCAAAGCCGCTGCTCCAAATGACATCCCAACAGTTGACACCAACCCGCT 398
OY 101 ProGluProSerGluThrAlaGluProThrAlaGluProThrGluGluProThrAlaGlu 120
DB 399 CCCGAGCCATCCGAGACCGCTGAGCCGCTGAGCCCAACGAGAGAGCCCACTGCGAG 458
OY 121 ProThrAlaGluProThrAlaGluProThrHisGluProThrGluGluProThrAlaVal 140
DB 459 CCTACCGCTGAGCCCAACCGCTGAGCTCATGAGCCCAACGAGAGAGCCCACTGCGAG 518
OY 141 ProThrGlyThrGlyGlyValProThrGlyThrGlySerPheThrValThrGlyArg 160
DB 519 CCACAGGACACTGGGGTGGTGTCCCACTGGACCGGTTCTTCAACGCTCATCGGAGA 578
OY 161 ProThrAlaSerThrProAlaGluPheProGlyAlaGlySerAsnValArgAla-SerVal 180
DB 579 CCAACTGGCTCCACCCAGCTGAGTTCCAGGTGCTGCTCCAAAGTCCGTCGCCAGCT 638
OY 180 TGYGlyYlleAlaAlaAlaLeuLeuGlyLeuAlaAlaIleValLeu 194
DB 639 TGGCGGCAATGCTGCTGCTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCTCGTCT 681

RESULT 2
BF252723

LOCUS BF252723 557 bp mRNA linear EST 15-NOV-2001
DEFINITION EST419986 Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAA21 5' sequence, mRNA sequence.

ACCESSION BF252723
VERSION BF252723.1 GI:16932866
KEYWORDS EST.
SOURCE Coccidioides immitis
ORGANISM Coccidioides immitis

REFERENCE 1 (bases 1 to 557)
AUTHORS Gardner, M.J. and Kirkland, T.
TITLE Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL Unpublished
COMMENT Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org.

FEATURES
source Location/Qualifiers
1..557

/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAA21"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 110 a 197 c 117 g 133 t
ORIGIN

Alignment Scores:
Pred. No.: 2.21e-45 Length: 557
Score: 706.00 Matches: 132
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 69.01% Indels: 0
DB: 10 Gaps: 0

US-10-081-935-4 (1-194) x BF252723 (1-557)

OY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
DB 161 ATGCAATTCTCTCACGCTCTCATCGCTCGTGGCGGCGCTCGCAAGTCCAGGCTC 220
OY 21 ProAapIleProProCysAlaLeuAsnCysPheValGluAlaLeuGlyAsnAspGlyCys 40
DB 221 CCAGACATCCCACTTGGCTCTCAACTGCTTGTGAGGCTCTCGCAACGATGGCTGC 280
OY 41 ThrArgLeuThrAspPheLysCysHisCysSerLysProGluLeuProGlyGlnIleThr 60
DB 281 ACTGCTTGACGCACTTCAAGTCCACTGCTCCAAAGCTGAGCTCCAGAGACAGATCACT 340
OY 61 ProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleValVal 80
DB 341 CCTGCGTTGAGAGAGCGCTGCCCTCTCGACGCCGATCTCGTCTCCAAACATCGCTT 400
OY 81 AspGlnCysSerLysAlaGlyValProIleAapIleProProValAspThrThrAlaAla 100
DB 401 GACCAAGTCTCCAAAGCCGCTGCTCCAAATGACATCCCAACAGTTGACACCAACCGCT 460
OY 101 ProGluProSerGluThrAlaGluProThrAlaGluProThrHisGluProThrGluGlu 120
DB 461 CCCGAGCCATCCGAGACCGCTGAGCCCAACCGCTGAGCCCAACGAGAGAGCCCACTGCGAG 520
OY 121 ProThrAlaGluProThrAlaGluProThrHisGlu 132

```

Db      521 CCTACCGCTGAGCCCAACCGCTGAGCCGACTCATGAG 556
|||||
RESULT 3
BF251334
LOCUS   EST418594 Coccidioides immitis spherule cDNA library Coccidioides
DEFINITION
immitis cDNA clone CIAAF68 5' sequence, mRNA sequence.
ACCESSION
BF251334
VERSION
BF251334.1 GI:16931477
KEYWORDS
EST.
SOURCE  Coccidioides immitis
        Coccidioides immitis
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
          Oryziales; microsporid Oryziales; Coccidioides.
REFERENCE
1 (bases 1 to 531)
AUTHORS Gardner,M.J. and Kirkland,T.
TITLE   Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL Unpublished
COMMENT Contact: Malcolm J. Gardner
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Drive, Rockville, MD 20850, USA
          Tel: 301 838 3519
          Fax: 301 838 0208
          Email: gardner@tigr.org.

FEATURES
        source
            1..531
                /organism="Coccidioides immitis"
                /mol_type="mRNA"
                /db_xref="taxon:5501"
                /clone="CIAAF68"
                /dev_stage="spherule"
                /lab_host="SOLR"
                /clone_1lb="Coccidioides immitis spherule cDNA library"
                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                XhoI"

BASE COUNT  107 a 190 c 115 g 119 t
ORIGIN
1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaIleArgIleuAlaSerAlaGlnLeu 20
|||||
Db      137 ATGCAATGTTCTCTCAAGCTCTCAATGCTCTGCTGCTGCTGCGCCCTCGCAGTGGCCACATC 186
|||||
Qy      21 ProAspIleProProCysAlaLeuAsnCysPheValGluAlaLeuGluAsnAspGlyCys 40
|||||
Db      197 CCAAGACATCCACATCTTGGCTCAACTGCTTGTAGAGCTCTCGGCAACGATGGCTGC 256
|||||
Qy      41 ThrArgLeuThrAspPheLysCysHisCysSerLysProGluLeuProGlyGlnIleThr 60
|||||
Db      257 ACTCGCTTGACCACTTCAGATGCACTGCTCCAAAGCTAGCTCCAGAGACAGATACT 316
|||||
Qy      61 ProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAlaIleValAla 80
|||||
Db      317 CCTTGGCTTGAGAGAGGCTCCCTCTCGACGCCGCTATCTCGCTCCAAATGCTGCTT 376
|||||
Qy      81 AspGlnCysSerLysAlaGlyValProIleAspIleProProValAspThrThrAlaAla 100
|||||
Db      377 GACCAATGCTCCAAAGCGGCTGCCAATTGACATCCACAGTTGACACACCGCGCT 436
|||||
Qy      101 ProGluProSerLysThrAlaGluProThrAlaGluProThrGluGluProThrAlaGlu 120
|||||
Db      437 CCGAGCCATCCAGACCGCTGAGCCCAACCGCTGAGCAACGAGAGGCCACTGCGCAG 496
|||||

```

```

Qy      121 ProThrAlaGluProThrAlaGluProThrHis 131
|||||
Db      497 CTAACCGCTGAGCCCAACCGCTGAGCCGACTCAT 529
|||||
RESULT 4
BF251817
LOCUS   EST419079 Coccidioides immitis spherule cDNA library Coccidioides
DEFINITION
immitis cDNA clone CIAAM80 5' sequence, mRNA sequence.
ACCESSION
BF251817
VERSION
BF251817.1 GI:16931960
KEYWORDS
EST.
SOURCE  Coccidioides immitis
        Coccidioides immitis
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
          Oryziales; microsporid Oryziales; Coccidioides.
REFERENCE
1 (bases 1 to 395)
AUTHORS Gardner,M.J. and Kirkland,T.
TITLE   Generation of ESTs from Coccidioides immitis spherule cDNA library
JOURNAL Unpublished
COMMENT Contact: Malcolm J. Gardner
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Drive, Rockville, MD 20850, USA
          Tel: 301 838 3519
          Fax: 301 838 0208
          Email: gardner@tigr.org.

FEATURES
        source
            1..395
                /organism="Coccidioides immitis"
                /mol_type="mRNA"
                /db_xref="taxon:5501"
                /clone="CIAAM80"
                /dev_stage="spherule"
                /lab_host="SOLR"
                /clone_1lb="Coccidioides immitis spherule cDNA library"
                /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                XhoI"

BASE COUNT  68 a 158 c 96 g 73 t
ORIGIN
1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaIleArgIleuAlaSerAlaGlnLeu 20
|||||
Qy      68 ProLeuAspAlaArgIleSerValSerAsnIleValAlaAspGlnCysSerLysAlaGly 87
|||||
Db      2 CTTCTGAGAGCCCGGTATCTCGCTCCAATCATGCTGTGACCAAGCTCTCCAAAGCGGCT 61
|||||
Qy      88 ValProIleAspIleProProValAspThrThrAlaAlaProGluProSerGluThrAla 107
|||||
Db      62 GTCCCAATTGACATCCACCAAGTTGACACCAACCGCCCTCCGAGCATCCGAGACCGCT 121
|||||
Qy      108 GluProThrAlaGluProThrGluGluProThrAlaGluProThrAlaGluProThrAla 127
|||||
Db      122 GAGCCCAACCGCTAGACCAACCGAGAGCCCACTGAGCTGACCGCTGAGCCCAACCGCT 181
|||||
Qy      128 GluProThrHisGluProThrGluGluProThrAlaValProThrGlyThrGlyGly 147
|||||
Db      182 GAGCCGACTCATGAGCCCAACCGAGAGCCCACTGTGCTCCAAACCGGCACTGGCGGTGCT 241
|||||
Qy      148 ValProThrGlyThrGlySerPheThrValThrGlyArgProThrAlaSerThrProAla 167
|||||
Db      242 GTCCCACTGGCAACCGGTTCTTACCGCTCATCGGACGAGCAACACTGCTCCACCCGAGCT 301
|||||
Qy      168 GluPheProGlyValaGlySerAsnValArgAlaSerValGlyGlyIleAlaIleAlaLeu 187
|||||

```

|||||
Db 302 GAGTTCAGGCTGCTGCTCCAGCTCCGTGCGACGTTGGCGCATTCGTGCTCTC 361
Qy 188 LeuGlyLeuAlaAlaTyrLeu 194
|||||
Db 362 CTCGGGCTCGCTGCTACCTG 382

RESULT 5
BF252856 401 bp mRNA linear EST 15-NOV-2001
LOCUS BF252856
DEFINITION Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAB70 5' sequence, mRNA sequence.
ACCESSION BF252856
VERSION BF252856.1 GI:16932999
KEYWORDS EST.
SOURCE Coccidioides immitis
ORGANISM Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.

REFERENCE
AUTHORS 1 (bases 1 to 401)
TITLE Gardner, M.J. and Kirkland, T.
JOURNAL Generation of ESTs from Coccidioides immitis spherule cDNA library
COMMENT Unpublished
Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@cigr.org.

FEATURES
source location/Qualifiers
1..401
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAB70"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 77 a 149 c 89 g 86 t
ORIGIN

Alignment Scores:
Pred. No.: 5.86e-38 Length: 401
Score: 608.00 Matches: 114
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 59.43% Indels: 0
DB: 10 Gaps: 0

US-10-081-935-4 (1-194) x BF252856 (1-401)

Qy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
Db 59 ATGAGATTCTCTACGCTCTCATCGCTCTGCTGCGGCTGCGAGTGCACAGTCTC 118
Qy 21 ProAspIleProProCysAlaLeuAsnCysPheValGluAlaLeuGlyAsnAspGlyCys 40
Db 119 CCAGACATCCACACCTTGCGCTCTCAACTGCTTCGTTGAGGCTCTGCGCAACGATGGCTGC 178
Qy 41 ThrArgLeuThrArgPheLeuPheLeuGlyCysHisCysSerLysProGluLeuProGlyGlnIleThr 60
Db 179 ACTGCTTGACCGACCTTCAAGTGCACCTGCTCCAAAGCTGAGCTCCCGACAGATTCAC 238
Qy 61 ProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleValVal 80
Db 239 CTTTGCGTTGAGGAGGCGCTGCTCGACCGCGGTATCTCGTCTCCACATCGTCTT 298
Qy 81 AspGlnCysSerLysAlaGlyValProIleAspIleProProValAspThrThAlaAla 100
Db 299 GACCAAGTCTCAAGGCGGCTGCTCCCAATTGACATCCACCAAGTTGACACACGCGCT 358

Qy 101 ProGluProSerGluThrAlaGluProThrAlaGluProThr 114
Db 359 CCCAGGCCATCCGAGACCGCTGAGCCACCGCTGAGCCAC 400

RESULT 6
BF251279 395 bp mRNA linear EST 15-NOV-2001
LOCUS BF251279
DEFINITION Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAAFl 5' sequence, mRNA sequence.
ACCESSION BF251279
VERSION BF251279.1 GI:16931422
KEYWORDS EST.
SOURCE Coccidioides immitis
ORGANISM Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.

REFERENCE
AUTHORS 1 (bases 1 to 395)
TITLE Gardner, M.J. and Kirkland, T.
JOURNAL Generation of ESTs from Coccidioides immitis spherule cDNA library
COMMENT Unpublished
Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@cigr.org.

FEATURES
source location/Qualifiers
1..395
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAAFl"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 78 a 133 c 78 g 106 t
ORIGIN

Alignment Scores:
Pred. No.: 8.03e-26 Length: 395
Score: 451.00 Matches: 84
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 44.09% Indels: 0
DB: 10 Gaps: 0

US-10-081-935-4 (1-194) x BF251279 (1-395)

Qy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
Db 142 ATGAGATTCTCTACGCTCTCATCGCTCTGCTGCGGCTGCGAGTGCACAGTCTC 201
Qy 21 ProAspIleProProCysAlaLeuAsnCysPheValGluAlaLeuGlyAsnAspGlyCys 40
Db 202 CCAGACATCCACACCTTGCGCTCTCAACTGCTTCGTTGAGGCTCTGCGCAACGATGGCTGC 261
Qy 41 ThrArgLeuThrArgPheLeuPheLeuGlyCysHisCysSerLysProGluLeuProGlyGlnIleThr 60
Db 262 ACTGCTTGACCGACCTTCAAGTGCACCTGCTCCAAAGCTGAGCTCCCGACAGATTCAC 321
Qy 61 ProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleValVal 80
Db 322 CTTTGCGTTGAGGAGGCGCTGCTCGACCGCGGTATCTCGTCTCCACATCGTCTT 381
Qy 81 AspGlnCysSer 84
Db 382 GACCAAGTCTCC 393

RESULT 7

LOCUS	BQ767421	602 bp	mRNA	linear	EST 26-JUL-2002
DEFINITION	EBR008_S0008_122_R root, 3 week, drought-stressed, cv Optic, EBR008 Hordeum vulgare subsp. vulgare cDNA clone EBR008_S0008_122 5', mRNA sequence.				
ACCESSION	BQ767421				
VERSION	BQ767421.1				
KEYWORDS	GI:21975895				
SOURCE	EST.				
ORGANISM	Hordeum vulgare subsp. vulgare				
	Hordeum vulgare subsp. vulgare				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae; Triticeae; Hordeum.				
REFERENCE	1 (bases 1 to 602)				
AUTHORS	Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L., Ramey, L., Mearns, G., Marshall, D.F.M. and Waugh, R.				
TITLE	Development of Barley Transcriptome Resources				
JOURNAL	Unpublished				
COMMENT	Contact: Waugh R, Marshall DF Genome Dynamics/Computational Biology Scottish Crop Research Institute Invergowrie, Dundee, DD2 5DA, Scotland, UK Tel: 00 44 1382 562731 Fax: 00 44 1382 562426 Email: eeb@scri.sari.ac.uk.				
FEATURES	Location/Qualifiers				
source	1..602				
	/organism="Hordeum vulgare subsp. vulgare"				
	/mol_type="mRNA"				
	/culivar="Optic"				
	/db_xref="taxon:112509"				
	/clone="EBR08_S0008_122"				
	/tissue_type="root"				
	/dev_stage="3 week"				
	/lab_host="DH10B"				
	/clone_lib="root, 3 week, drought-stressed, cv Optic, EBR08"				
	/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from roots of 3 week old drought stressed barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project."				
BASE COUNT	100 a 251 c 129 g 122 t				
ORIGIN					
Alignment Scores:					
Prod. No.:	1.22e-22	Length:	602		
Score:	412.50	Matches:	84		
Percent Similarity:	59.8%	Conservative:	25		
Best Local Similarity:	46.1%	Mismatches:	56		
Query Match:	40.3%	Indels:	17		
DB:	13	Gaps:	1		
US-10-081-935-4 (1-194) x BQ767421 (1-602)					
Oy	1	MecGlnhseSRH:sAlaLeuAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu	20		
Db	45	ATGAAGTTCTCCACCACTCTGATTCCTTCGTCGGCGCTCGGCGACGCCACTG	104		
Oy	21	ProAspIleProProCySAIalaLeuAsnCysPheValGluAlaLeuGlyAsnAspGlyCys	40		
Db	105	CCCGATGCGCCGCTGCTGCTGAGCTGCTTCCTCTGCTTCGCTTGACAGATGCTTGC	164		
Oy	41	ThrArgLeuThrAspPheIysCysHisCysSerIysProGluLeuProGlyGlnIleThr	60		
Db	165	TCCAGCTGCTGCCTTCACTGCACCTGCAGAGCCGAGCTCGTAGCAGCATACT	224		
Oy	61	ProCySAIgluGluAlaCySProlLeuAspAlaArgIleSerValSerAsnIleValVal	80		
Db	225	CCTTGTGTTCGAAGAGCTGCGAGTCAAGACCAATTCCTCCGTCACAGCCCTGCTC	284		
Oy	81	AspGlnCysSerIysAlaGlyValAlaProIleAspIleProProValAspThrThrAlaAla	100		

Db	285	GGCAGTGTCTTTCGCCGGCCGACCCATCTCGATTCCTCCCGCATGAGACCAAGCCGCTTC	344
Qy	101	ProgluPProsergluThralagluProthralagluProthrgluProthralaglu	120
Db	345	TCGAGCGCTCCAGAGCCTCCAGACCTCGAGCTCCGCCCGCTGAGACCAAGCCCGAG	404
Qy	121	ProthralagluProthralagluProthralagluProthrgluProthrgluProthrala	140
Db	405	ACCAACCGCTCCAGCAGGAGAGCTCCAGCCCGTTGAGTCTCTCTCTCTCCGCTGT	464
Qy	141	Prothrgly	143
Db	465	GAGACGGGACCTCCACCAACCGTCCGAGACTACTACCGCTGTTCCACCAACGTC	524
Qy	144	ThrglyglyyvalProthrglyThrglyserphethrvalThrglyargProthrgla	163
Db	525	AGCTCTCCGGTACTGCCACTCCACCCGCTGAGCAGCACTACCGCTCCAGCACTGCC	584
Qy	164	Se7Thr 165	
Db	585	TCTACC 590	
RESULT 8			
LOCUS	BQ767128	538 bp	mRNA linear EST 26-JUL-2002
DEFINITION	EBro08 S0008 E05 R root, 3 week, drought-stressed, cv Optic, EBro0808 Hordeum vulgare subsp. vulgare cDNA clone EBro08_S0008_E05 5', mRNA sequence.		
ACCESSION	BQ767128		
VERSION	BQ767128.1	GI:21975602	
KEYWORDS			
SOURCE	Hordeum vulgare subsp. vulgare		
ORGANISM	Hordeum vulgare subsp. vulgare		
REFERENCE	Healey, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L., Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.		
AUTHORS	Development of Barley Transcriptome Resources		
TITLE	Unpublished		
JOURNAL			
COMMENT	Contact: Waugh R, Marshall DP Genome Dynamics/Computational Biology Scottish Crop Research Institute Invergowrie, Dundee, DD2 5DA, Scotland, UK Tel: 00 44 1382 562731 Fax: 00 44 1382 562426 Email: est@scri.sari.ac.uk.		
FEATURES			
source	Location/Qualifiers		
	1..538		
	/organism="Hordeum vulgare subsp. vulgare"		
	/mol_type="mRNA"		
	/culivar="Optic"		
	/db_xref="taxon:112509"		
	/clone="EBro08_S0008_E05"		
	/classue_type="root"		
	/dev_stage="3 week"		
	/lab_host="DH10B"		
	/clone_id="root, 3 week, drought-stressed, cv Optic, EBro08"		
	/note="Vector: pSPORT1, Site_1: Sal I, Site_2: Not I. Non-normalised library, directionally cloned into pSPORT1. Derived from roots of 3 week old drought stressed barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project."		
BASE COUNT	91 a 217 c 119 g 111 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1.09e-21	length:	538
Score:	399.50	Matches:	80

Percent Similarity: 64.67%
Best Local Similarity: 47.90%
Query Match: 39.05%
DB: 13
Conservative: 28
Mismatches: 48
Indels: 11
Gaps: 2

US-10-081-935-4 (1-194) x BQ761304 (1-538)

```
Qy      1 MetGlnPheSerHsAlaLeuAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      56 ATGAAGTTCTCCAGCCTGATGATGCTTCCTGCTCCCGCGGCTCCGAGCGCCAGCTG 115
Qy      21 ProAspIleProProCysAlaLeuAsnCysPheValGluAlaLeuGlyAsnAspGlyCys 40
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      116 CCCATGTGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 175
Qy      41 ThrArgLeuThrAspPheLysCysHisCysSerIleProGluLeuProGlyGlnIleThr 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      176 TCCAGAGCTCTCCAGCCTTCACTGACCTCCAGAACCCGAGCTGCTGAGAGCATCACT 235
Qy      61 ProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleValAl 80
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      236 CTTGTGTTCCAGAGCCTCCAGGTCAGAGCCAAATCTCCGTCCTCAACGCCGCTCTC 295
Qy      81 AspGlnCysSerLysAlaGlyValProIleAspIleProProValAspThrThrAlaAla 100
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      296 GCGCAGTGTCTTCCGCGCGCCACCCATCTGATTCCTCCCATCGAGACAGCGCTTCC 355
Qy      101 ProGluProSerGluThrAlaGluProThrAlaGluProThrGluGluProThrAlaGlu 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      356 TCCAGCGCGCTCCAGAGCCTCCAGACCTCCGAGCTCCGCGCGCGGAGACAGCGCGAG 415
Qy      121 ProThrAlaGluProThrAlaGluProThrIleGluProThrGluGluProThrAlaVal 140
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      416 ACCACCGCGC-----TCCACACCGAGAGCTCCAGCGCGCTT 451
Qy      141 ProThrGlyThrGlyGlyValProThrGlyThrGlySerPheThrValThrGlyArg 160
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      452 GAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 502
Qy      161 ProThrAlaSerThrProAla 167
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      503 GCAGTACTACTACCGCTGCT 523
RESULT 9
BQ761304 419 bp mRNA linear EST 26-JUL-2002
LOCUS EB006 SQ001 A05 R root, 3 week, drought-stressed, cv Optic, EB006
DEFINITION Hordeum vulgare subsp. vulgare cDNA clone EB006_SQ001_A05 5', mRNA
sequence.
ACCESSION BQ761304
VERSION BQ761304.1 GI:21969776
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
REFERENCE 1 Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardie,L.,
AUTHORS Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished
COMMENT Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk.
Location/Qualifiers
FEATURES
source 1..419
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
```

```
/cultivar="Optic"
/db_xref="taxon:112509"
/clone="EB006_SQ001_A05"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="root, 3 week, drought-stressed, cv Optic,
EB006"
/notes="vector: pSPORT1, Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from roots of 3 week old drought stressed barley
plants. After two weeks in control conditions the seedlings
were raised above the hydroponic solution for 3 hours
daily. N.B. This library has been superseded by EB008.
Developed as part of the barley transcriptome resources of
BBRC/SEERAD funded cereal IGF (Investigating Gene
Function) project."
BASE COUNT 71 a 169 c 97 g 82 t
ORIGIN
Alignment Scores:
Pred. No.: 1.86e-20 Length: 419
Score: 382.00 Matches: 73
Percent Similarity: 67.69% Conservative: 15
Best Local Similarity: 56.15% Mismatches: 41
Query Match: 37.34% Indels: 1
DB: 13 Gaps: 0
US-10-081-935-4 (1-194) x BQ761304 (1-419)
Qy      1 MetGlnPheSerHsAlaLeuAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      27 ATGAAGTTCTCCAGCCTGATGATGCTTCCTGCTCCCGCGGCTCCGAGCGCCAGCTG 86
Qy      21 ProAspIleProProCysAlaLeuAsnCysPheValGluAlaLeuGlyAsnAspGlyCys 40
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      87 CCCATGTGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 146
Qy      41 ThrArgLeuThrAspPheLysCysHisCysSerIleProGluLeuProGlyGlnIleThr 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      147 TCCAGAGCTCTCCAGCCTTCACTGACCTCCAGAACCCGAGCTGCTGAGAGCATCACT 206
Qy      61 ProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleValAl 80
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      207 CTTGTGTTCCAGAGCCTCCAGGTCAGAGCCAAATCTCCGTCCTCAACGCCGCTCTC 266
Qy      81 AspGlnCysSerLysAlaGlyValProIleAspIleProProValAspThrThrAlaAla 100
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      267 GCGCAGTGTCTTCCGCGCGCCACCCATCTGATTCCTCCCATCGAGACAGCGCTTCC 326
Qy      101 ProGluProSerGluThrAlaGluProThrAlaGluProThrGluGluProThrAlaGlu 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      327 TCCAGCGCGCTCCAGGCTCCAGACCTCCGAGCTCCGCGCGCGGAGACCGCGAG 386
Qy      120 uProThrAlaGluProThrAlaGluPro 129
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      387 ACCACCGCTCTCCAGCAGGAGACTCCA 414
RESULT 10
BQ499065 311 bp mRNA linear EST 31-OCT-2002
LOCUS EST08290 Pb0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.
DEFINITION BQ499065
ACCESSION BQ499065
VERSION BQ499065.1 GI:24452039
KEYWORDS EST.
SOURCE Paracoccidioides brasiliensis
ORGANISM Paracoccidioides brasiliensis
REFERENCE 1 Omygenales; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
AUTHORS Goldman,G.H., Marques,E.R., Ribeiro,D.C.D., Bernardes,L.A., Puccia
Goldman,G.H., Marques,E.R., Ribeiro,D.C.D., Bernardes,L.A., Puccia
,R., Travassos,L.R., Nobrega,M.P., Savoldi-Barbosa,M.
,Semighini,C.P. and Goldman,M.H.
```

JOURNAL COMMENT		TITLE	
The Paracoccidioides brasiliensis EST genome project		Eukaryotic Cell, (2002) In press	
Contact: Gustavo Henrique Goldman		Laboratory of Molecular Biology	
Universidade de Sao Paulo - USP - FCRP		Av do Cafe S/N, CEP: 14040-903, Ribeirao Preto - SP, Brazil	
Email: ggoldman@usp.br			
FEATURES		Location/Qualifiers	
source		1..311	
BASE COUNT	73 a 88 c 59 g 91 t		
ORIGIN	/organism="Paracoccidioides brasiliensis" /mol_type="mRNA" /db_xref="taxon:121759" /clone_1fb="Pb0001"		
Alignment Scores:			
Pred. No.:	5,87e-16	Length:	311
Score:	322.00	Matches:	56
Percent Similarity:	93.06%	Conservative:	11
Best Local Similarity:	77.78%	Mismatches:	5
Query Match:	31.48%	Indels:	0
DB:	13	Gaps:	0
US-10-081-935-4 (1-194) x BQ499065 (1-311)			
OY	1	MecGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu	20
DB	91	ATGCAGTCTCTCCAGCGCTCTCATCTGTTGTGGCGCCAGCTTACCAATCTCAGCTC	150
OY	21	ProbaGllleProProCysAlaLeuAnsCysPheValGluAlaLeuGlyAnsAspGlyCys	40
DB	151	CCTGATATCCACACTTGGCGTTTGAGCTGTTTCGTGATCTCTCGAAACGACGGGTGC	210
OY	41	TharGLeuThrAspPheLysCysHisCysSerLysProGluLeuProGlyGlnIleThr	60
DB	211	TCCACGCTGACCGAGATTGCGAGTCCACATCGCAAAAAGCTGAGTTCGCCGAAAGATCACT	270
OY	61	ProCysValGluGluAlaCysProLeuAnsAlaArg	72
DB	271	CCTTGGTTGAGAAATCTTGACGCGTCTCGCCCA	306
RESULT 11			
LOCUS	BM077163	667 bp	mRNA linear EST 05-FEB-2002
DEFINITION	TrEST-A0723 TrEST-A Hypocrea jecorina cDNA clone Tr-A0723 5', mRNA sequence.		
ACCESSION	BM077163		
VERSION	BM077163.1	GI:18498345	
KEYWORDS	EST.		
SOURCE	Hypocrea jecorina		
ORGANISM	Hypocrea jecorina		
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.		
AUTHORS	1 (bases 1 to 667) Chambergo, F.S., Bonaccorsi, E.D., Ferreira, A.J.S., Ramos, A.S.P., Ferreira, J.R.Jr., Abrahamo-Neto, J., Farah, J.P.S. and El-Dorri, H.		
TITLE	Elucidation of the metabolic fate of glucose in the filamentous fungus Trichoderma reesei using expressed sequence tag (EST) analysis and cDNA microarrays		
JOURNAL	J. Biol. Chem. 277 (16), 13983-13988 (2002)		
MEDLINE	21950703		
PUBMED	11825887		
COMMENT	Contact: El-Dorri, Hamza Department of Biochemistry Institute of Chemistry, University of Sao Paulo Avenida Professor Lineu Prestes, 748, Sao Paulo, SP, 05508-900, BRASIL Tel: (55) 11-38183848 Fax: (55) 11-38183848 Email: dorri@iq.usp.br PCR Primers FORWARD: Universal M13 forward primer		

FEATURES	
SOURCE	1..667 Location/Qualifiers
BACKWARD: Universal M13 reverse primer Plate: 8 row: E column: 3 Seq primer: M13 reverse primer High quality sequence stop: 667 POLYA=No.	
/organism="Hypocrea jecorina" /mol_type="mRNA" /strain="QM9414 (ATCC26921)" /db_xref="taxon:51453" /clone="Tr-A0723" /sex="Asexual" /tissue_type="Mycelia" /dev_stage="18 hr glycerol-grown culture" /lab_host="E. coli S0LR cells (kanamycin resistant)" /clone_lib="TREST-A" /note="Vector: pBlueScript SK(+); Site 1: EcoRI; Site 2: XhoI; anamorph=Trichoderma reesei; Cloned unidirectional , 5' end of the cDNA cloned into EcoRI site of pBlueScript. Primer: Oligo (dT). Average insert size: 1, kb. Uni-ZAP XR Vector system -5' adaptor sequence: 5'GAATTGGCAGCAGG3' -3' adaptor sequence: 5'CTCAGGCTTTTCTTTTCTTTT3'"	
BASE COUNT	98 a 266 c 188 g 115 t
ORIGIN	
Alignment Scores:	
Pred. No.:	1.16e-07 Length: 667
Score:	219.50 Matches: 72
Percent Similarity:	44.19% Conservative: 23
Best Local Similarity:	33.49% Mismatches: 69
Query Match:	21.46% Indels: 51
DB:	12 Gaps: 14
US-10-081-935-4 (1-194) x BM077163 (1-667)	
QY	23 11eProProCySa1a1eua5ncYspheVal1G1ua1a1eulG1yA5na5pJ1CyethrArg 42
DB	33 GTCCCTCATGCGCCCAAGCGTGCTCTCTGCGCGCCCGGGCGCTGCGTCCAC--- 89
QY	43 1euThrAspPhelYscYeh1scYser1yPProGluLeuProGlyGln1leThr----- 60
DB	90 GTCMAAGACTTGACAGCCAGTGC---AAGCCGAGGCTCAAG:CCAGGCTGACCATCTG 14
QY	61 -----ProCySa1a1GluG1ua1a1CyPro-----1euAsp1a1Arg1leSerVal 75
DB	147 CTGAGTCCCTGCGTCCGACCTGCTGCTCTGCTGCTCTCTGACAGCGGTATCAACCGCA 20
QY	76 SerAsn11eVal1a1a5pG1ncYser1ySa1a1yValPro1leAsp11eProProVal 95
DB	207 GCCAGCTCAGTGC---GCTGTGCGACGCGCTGCGCCACGCTGTTGCCCTGGACT 26
QY	96 AspThrThAla1a1AProGluPro-----SerGluThrAlaGlu---ProThr 111
DB	264 TGCAGACCGAGAGCCCTCTCCGCCCGCGGACGACTCAAGACCCGGAAGTCCCCAGC 32
QY	111 AlaGluProThG1uG1uProThrAlaGluPro-----ThrAla 122
DB	324 AGCTGCCGCGTACGACGCCCTCCGGAAGCCGCTCTCTGCTTCTTCTTCTCTCT 38
QY	124 GluProThrAlaGluProThrThiGluProThr----- 133
DB	384 CCCCCGAGCAGCAGCTTACCGAGACCCCACTGCGCGCAAGGAGAACGACCGCTCT 44
QY	135 -----GluGluProThrAla-----ValProThG1yThhG1y 144
DB	444 GGACCAAGGCGGAACAGCCACCGAGGATTCACACACATGCCCCAGGCGCTT---GGC 500
QY	147 G1yValProThG1yThhG1ySerPherThValThhG1yArgProThrAlaSerThPro 166
DB	501 TCGACGCGACCGGACCTGGTTCCGAGACGACTGAGAGAGAGAGCGGAAGCCAGCC 566

QY 167 AlaGlu-----PheProGlyAlaGlySerAsnValArgAlaSerVal----- 180
 Db 561 ACTGGCGCCGACGACTCTTCCCCGAGCGCGCTTCTCCAGTCGAGATGACGCTTGCT 620
 QY 181 GLyGlyIleAla-----AlaAlaLeuLeuGlyLeu 190
 Db 621 TGTGCTTTTGGCGCTACTGCGGCAATTCAGTTTCTGCGGATTC 665

RESULT 12
 BM66652 564 bp mRNA linear EST 06-MAY-2003
 LOCUS mgcs006x016f.b Magnaporthe grisea CS Uni-Zap XR Library Magnaporthe
 DEFINITION grisea cDNA clone mgcs006x016 5', mRNA sequence.
 ACCESSION BM66652.1 GI:19233334
 VERSION 1
 KEYWORDS Magnaporthe grisea (anamorph: Pyricularia grisea)
 SOURCE Magnaporthe grisea
 ORGANISM Magnaporthe grisea (anamorph: Pyricularia grisea)
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
 1 (bases 1 to 564)
 Ebbole,D.J., Yuan,J., Thomas,T.L., Bobrowicz,P., Lu,G., Bhatterai
 K. and Dean,R.A.
 Expressed sequence tags from the rice blast fungus, Magnaporthe
 grisea
 TITLE Unpublished
 CONTACT: Ebbole DJ
 DEPARTMENT of Plant Pathology & Microbiology
 Texas A&M University
 Peterson Bldg, MS2132, College Station, TX 77843-2132, USA
 Tel: 979 845 4831
 Fax: 979 845 6483
 Email: d-ebbole@tamu.edu
 Chromatogram file of this sequence is available, see contact person
 [Best nr hit (April. 22, 2003) gb|EAA29323.1| predicted protein
 (Neurospora crassa) 55 6e-07
 PCR primers
 FORWARD: T3 primer
 BACKWARD: T7 primer
 Plate: mgcs006 row: B column: 09
 Seq primer: T3.

FEATURES
 source 1..564
 Location/Qualifiers
 /organism="Magnaporthe grisea"
 /mol_type="mRNA"
 /strain="Guy11"
 /db_xref="taxon:148305"
 /clone="mgcs006x016"
 /sex="Mat1-2 hermaphrodite"
 /cell_type="conidia"
 /clone_1fb="Magnaporthe grisea CS Uni-Zap XR Library"
 /note="Vector: pBluescriptSK-; Site_1: EcoRI; Site_2: XhoI
 ; Unidirectional cloning. EcoRI side has T3 primer and
 predominantly 5' reads. T7 primer on XhoI side of insert.
 Conidial library. Point inoculation of Guy11 at center of
 oatmeal agar plate. Conidia were harvested after two weeks
 of growth. Sequences were processed by the M. grisea genome
 sequence was available, the EST sequence was trimmed
 according to the alignment, otherwise sequence quality was
 assessed using phredPhrap version 99109 and trimmed
 according to phd files (0.05) and for vector segs."

BASE COUNT 100 a 242 c 128 g 94 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,98e-06 Length: 564
 Score: 202.50 Matches: 64
 Percent Similarity: 47.42% Conservative: 28
 Best Local Similarity: 32.95% Mismatches: 81
 Query Match: 19.79% Indels: 22
 DB: 12 Gaps: 10

US-10-081-935-4 (1-194) x BM66652 (1-564)
 QY 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerValaGlnLeu 20
 Db 33 ATGAAGTTCTCC---ATCATCTCGCGCGCGCGCGCACGCGCTGCTCGACGACTTC 89
 QY 21 ProAspIleProProGlyAlaLeuAsnGlyPheValGluAlaLeuGlyAsnAspGlyCys 40
 Db 90 TCTTCCCTCCCGCGCTGCGTGTGACATGTGTGACACGCGCTATTGCTTCCGCTGC 149
 QY 41 ThrArgLeuThrAspPheLeuGlySerHisGly-----SerLeuProGluLeuProGlyGln 58
 Db 150 AACGCG---ACCGACGTGCGCTCCGACGTGTGTGAGAGAGAGAGACATTGCCAAGTC 206
 QY 59 IleThrProGlyValGluGluAlaGlyProLeuAspAlaArgIleSerValSerAsnIle 78
 Db 207 GCTACTCTTCTGCTGCTGCGACGACGACGACCTGCGACCTCTCCAGGCTCCACC 266
 QY 79 -ValValAspGlnGlySerLeuValAlaGlyValProIleAspIleProValAspThrTh 98
 Db 267 AACGCGAGCGCGCTGTGAGAACGTCGCCACCGCTGTTCTTACTCTCT-----CC 317
 QY 98 AlaAlaProGluProSerGluThrAlaGluProThrAlaGluProThrGluGluProTh 118
 Db 318 AGCGCGCGCTTACCACCAACCCACC---CCGCAACTTTCGCGCTCGCGCGCACGACC 374
 QY 118 AlaGluProThrAlaGluProThrAlaGluProThrHisGluProThrGluGluProTh 138
 Db 375 ATGCCCTCATGTGATGATGCGTGTGCGCCCAAGCAACGCG-ACCTGCTGCTCTCCACC 433
 QY 138 AlaValProThrGluThrGlyGlyGlyVal-----ProThrGlyThrGlySerPheTh 156
 Db 434 CGCTTCCCGCCACGACGACGCGGAGAGAGAACCCGCGGACGACGCGC----- 488
 QY 156 rValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyAlaGlySerAsnVa 176
 Db 489 -----GGCTCTCCCTCTGCGACGACGCTGCC-----GGCTCTGCGCGCTCCGC 532
 QY 176 lArgAlaSerValGlyGlyIleAlaAlaLeuLeuGly 189
 Db 533 C---GCCACTATCGCG-----GCTGCTGCTCTTCTTGST 563

RESULT 13
 BM666040 553 bp mRNA linear EST 06-MAY-2003
 LOCUS mgcs004x016f.b Magnaporthe grisea CS Uni-Zap XR Library Magnaporthe
 DEFINITION grisea cDNA clone mgcs004x016 5', mRNA sequence.
 ACCESSION BM666040
 VERSION 1
 KEYWORDS Magnaporthe grisea (anamorph: Pyricularia grisea)
 SOURCE Magnaporthe grisea
 ORGANISM Magnaporthe grisea (anamorph: Pyricularia grisea)
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
 1 (bases 1 to 553)
 Ebbole,D.J., Yuan,J., Thomas,T.L., Bobrowicz,P., Lu,G., Bhatterai
 K. and Dean,R.A.
 Expressed sequence tags from the rice blast fungus, Magnaporthe
 grisea
 TITLE Unpublished
 CONTACT: Ebbole DJ
 DEPARTMENT of Plant Pathology & Microbiology
 Texas A&M University
 Peterson Bldg, MS2132, College Station, TX 77843-2132, USA
 Tel: 979 845 4831
 Fax: 979 845 6483
 Email: d-ebbole@tamu.edu
 Chromatogram file of this sequence is available, see contact person
 [Best nr hit (April. 22, 2003) gb|EAA29323.1| predicted protein
 (Neurospora crassa) 51 8e-06
 PCR primers
 FORWARD: T3 primer
 BACKWARD: T7 primer

Plate: mgcs004 row: 0 column: 16
 Seq primer: T3.
 Location/Qualifiers
 1.553

FEATURES
 source
 /organism="Magnaporthe grisea"
 /mol_type="mRNA"
 /strain="Guy11"
 /db_xref="taxon:148305"
 /clone="mgcs004X016"
 /sex="Mat1-2 hermaphrodite"
 /cell_type="conidia"
 /clone_lib="Magnaporthe grisea CS Uni-Zap XR Library"
 /note="Vector: pBluescriptSK-; Site 1: EcoRI; Site 2: XhoI
 ; Unidirectional cloning. EcoRI site has T3 primer and
 predominantly 5' reads. T7 primer on XhoI site of insert.
 Conidia library. Point inoculation of Guy11 at center of
 oatmeal agar plate. Conidia were harvested after two weeks
 of growth. Sequences were processed by one of two methods.
 Where a full-length alignment to the M. grisea genome
 sequence was available, the EST sequence was trimmed
 according to the alignment, otherwise sequence quality was
 assessed using phredphrap version 991019 and trimmed
 according to phd files (0.05) and for vector seqs."
 BASE COUNT 83 a 233 c 132 g 105 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.05e-05 Length: 553
 Score: 193.00 Matches: 61
 Percent Similarity: 46.704 Conservative: 24
 Best Local Similarity: 33.524 Mismatches: 77
 Query Match: 18.874 Indels: 21
 DB: 12 Gaps: 9

US-10-081-935-4 (1-194) x BM86040 (1-553)

14 GtyleuAlaSerAlaGluProAAspIleProProCysAlaLeuAsnCysPheValGlu 33
 6 GGCCGTCGGCTCAGAGACCTCTCTCCCGCGCTGCGCTGCACGTCTGACACAC 65
 34 AAlaLeuGlyAsnAspGlyCysThrArgLeuThrAspPheLysCysHicys-----Ser 51
 66 GCTATTGCTCTTCGCGCTGCAACGCGC--ACCGAGCTGCGCTGCGCAGTGTGTGAAG 122
 52 LysProGluLeuProGlyGlnIleThrProCysValGluGluAlaCysProLeuAspAla 71
 123 AAGAAGACATTGCCAACCTCGTACTCTTCTGCTCTGCGCAAGTGCACCGACCTGGC 182
 72 ArgIleSerValSerAsnIle-ValValAspGlnCysSerLysAlaGlyValProIleAs 91
 183 GACCTCTCCAGAGCTGCCACCAACGCGAGCGCTGTGCAAGACGTCCCAACCGCTGT 242
 91 pIleProProValAspThrThrAlaAlaProGluProSerGluThrAlaGluProThrAl 111
 243 CTTACTCTCT-----CCAGCGCGCGCTTACCAACCAACGCGC--CCGCAACTCT 290
 111 AgluProThrGluGluProThrAlaGluProThrAlaGluProThrAlaGluProThrAl 131
 291 TCGGCTCGGCGGACGACCAACATGCCCTCATGTGCATGCGGTGCCCAAGACCAACGCG 350
 131 sGluProThrGluGluProThrAlaValProThrGlyThrGlyGlyVal-----Pr 149
 351 -ACCTCTGCTGCTCTCCACCGCTTCCCGCACGACGACGAGGCGGAGGAGGAGAACCC 409
 149 oThrGlyThrGlySerPheThrValThrGlyArgProThrAlaSerThrProAlaGluPh 169
 410 CAGCGGACCAACGCGC-----GGCTCTCTCTCTGCGCACCGCCCTGCGC----- 452
 169 eProGlyAlaGlySerAsnValAlaAlaSerValGlyGlyIleAlaAlaLeuLeuGlu 189
 453 -----GGCTCTGCGCGCTCGCGC-----GCCACTATCGC-----GCTGCTCTTCTTGG 499
 189 yLeu 190

Db 500 TCTT 503

RESULT 14
 CD458295/c 952 bp mRNA linear EST 03-JUN-2003
 LOCUS
 DEFINITION
 Fg08_09f09 R Fg08 AAFc ECORC Fusarium graminearum complex substrate
 CD458295
 Giberella zeae cDNA clone Fg08_09f09, mRNA sequence.
 ACCESSION
 VERSION
 CD458295.1 GI:31373035
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 1 (bases 1 to 952)
 Watson,R.J., Heyes,R., Chapados,J., Courroux,P., Harris,L.J., Hattori,
 J., Lacroix,C., Ouellet,T., Robert,L.S., Singh,J.A., Sprott,D. and
 Tinker,N.A.

TITLE
 A cDNA library prepared from Fusarium graminearum grown on a
 complex plant substrate
 JOURNAL
 Unpublished

COMMENT
 Contact: Watson, Robert.J.
 Eastern Cereal and Oilseed Research Centre
 Agriculture and Agri-Food Canada
 Bldg. 20, Central Experimental Farm, Ottawa, Ontario, KIA 0C6,
 CANADA
 Tel: (613) 759-1655
 Fax: (613) 759-1701
 Email: watsonrj@agr.gc.ca.
 Location/Qualifiers

FEATURES
 source
 1.952
 /organism="Giberella zeae"
 /mol_type="mRNA"
 /strain="DAOM 180378"
 /db_xref="taxon:5518"
 /clone="Fg08_09f09"
 /tissue_type="mycelium"
 /dev_stage="Asexual"
 /lab_host="E. coli DH10B"
 /clone_lib="Fg08_AAFc_ECORC_Fusarium_graminearum_complex_s
 ubstrate"
 /note="Vector: pBluescript II+; Site 1: EcoRI, Site 2:
 XhoI; Fusarium graminearum grown on a complex plant
 substrate-- wheat leaves treated to remove most of the low
 molecular weight, water-soluble components."
 BASE COUNT 222 a 211 c 288 g 226 t
 ORIGIN

Alignment Scores:
 Pred. No.: 7.98e-05 Length: 952
 Score: 185.00 Matches: 70
 Percent Similarity: 41.784 Conservative: 19
 Best Local Similarity: 32.864 Mismatches: 89
 Query Match: 18.084 Indels: 36
 DB: 14 Gaps: 9

US-10-081-935-4 (1-194) x CD458295 (1-952)

6 AAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGln---LeuProAAspIlePro 24
 833 GCTACCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 774
 25 ProCysAlaLeuAsnCysPheValGluAlaLeuGlyAsnAspGlyCysThrArgLeuThr 44
 773 GAGTGTGCTGCTCAATGCTTAAAGAAATTCATCCCAAGCGTGGCTGC---GAGCTGCAC 717
 45 AspPheLysCysHicysSer-----LysProGluLeuProGlyGlnIleThrPro 61
 716 AACACCGGATGCAATGCGAGGCTGTTCCAAACAAAGCTGCTCCCATCATCTCTCC 657
 62 CysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleValValAsp 81

Db	656	TGCTTGAACCGAAGCTTGCCAGGTGATGACCTACTCAAGGCCAGAAAGTGGCCGCCAG	597
Qy	82	GInCys-----SerLYsAlaGlyValProIleasp	91
Db	596	GCCCTGCAAGGCTATGCTGCCACCGCTGACTCCGGCTCCGCCACTGCCACTGCCCTCCACCAAG	537
Qy	92	-----IleProValaspThrThrAlaIleProGluProSerclunthr	106
Db	536	AGCTCGGCTCTGTTACCGTCAGCATCACAACCTTATCACTGCTCGCGTAAGCTTCCC	477
Qy	107	AlaGluProThrAlaGluProThrGluGluProThrAlaGluProThrAlaGlu-Proth	126
Db	476	GCCATCTTCTCAAGCATCCGTGATGAGAAGCCATCCTCGTTACTCCCGTCCGGC	417
Qy	126	rAlaGluProThrHisGluPro-----ThrGluProThrAlaIva	140
Db	416	AACCGAACCAATGACCAACAGACCTAGCTAGAGAGGTGATCCGGAGAGGC-GCTGTAC	358
Qy	140	lProThrGlyThrGlyGlyGlyValProThrGlyThrGlySerPheThrValThrGlyAr	160
Db	357	ACCACGGGAAACCTCTGGCGGC-----GGCGGGGACGGCACACATGGCGGTGC	307
Qy	160	gProThrAlaSerThrProAlaGluPheProGlyValaGlySerAsnValAlaGlaSerVa	180
Db	306	TAGCAGTGTCCCTACCGACGCT-----GGTGTGTACTGCTGGCCCTGTCGCAT	256
Qy	180	lGly-----GlyIleAlaAlaAlaLeuGluGlyLeu	190
Db	255	GGGTCTCTCGCATTATTCGTCGCTATCCCTCTG	219
RESULT 15			
BU066219		731 bp	mRNA
LOCUS			linear
DEFINITION			EST 26-AUG-2002
			For 9 K19_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA
ACCESSION			BU066219
VERSION			BU066219.1
KEYWORDS			GI:22506508
SOURCE			EST.
ORGANISM			Gibberella zeae
			Gibberella zeae
			Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
			Hypocryotaceae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE			1 (bases 1 to 731)
AUTHORS			Trail, F., Xu, J.-R., San Miguel, P., Halgren, R.G. and Kistler, H.C.
TITLE			Analysis of expressed sequence tags from Gibberella zeae (anamorph
			Fusarium graminearum)
JOURNAL			Fungal Genet. Biol. 38 (2), 187-197 (2003)
MEDLINE			22508120
PUBMED			12620255
COMMENT			Contact: Frances Trail Department of Plant Biology Michigan State University East Lansing, MI 48824, USA Tel: 517 432 2939 Fax: 517 353 1926 Email: trail@msu.edu
FEATURES			Plate: 9 row: K column: 19.
source			Location/Qualifiers
			1..731
			/organism="Gibberella zeae"
			/mol_type="mRNA"
			/strain="NRRL 31084"
			/db_xref="taxon:5518"
			/clone_id="Nitrogen-starved mycelia"
			/note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT			151 a 255 c 172 g 153 t
ORIGIN			
Alignment Scores:			
Pred. No.:			7,12e-05
Score:			184.00
Percent Similarity:			41.51%
Best Local Similarity:			32.08%
			Length: 731
			Matches: 68
			Conservative: 20
			Mismatches: 91

Query Match: 17.99% Indels: 34
DB: 13 Gaps: 8

US-10-081-935-4 (1-194) x BU066219 (1-731)

Oy 6 A1aleu1lea1leuValA1aa1ag1yleua1aseRa1agin---LeuProAsp1IePro 24
||| :||| ||||||||| ||||| |||
Db 120 GCTACCCCTTCCTCCTCGGCTGCAGCGGCTGGCTGTCTGCTCAGCACTCAGTGGCTAACACC 179
Oy 25 ProCyba1aleuAnaCySpheVa1GIua1aleuGIyVehnsbpCIyCvStrArgLeuThr 44
||||| ||||| :|||
Db 180 GAGTGCTCTCCAATGTCTTAAGAAATTGATCCCCAAGCTGGCTCC--GA CCTGCAC 236
Oy 45 AspheUeyCyShIScySer-----LysProGluLeuProGIyGlnIleThPro 61
::: ||||| ::| :|||
Db 237 AACACCGCATGCCATGCCAGGCTCGTTCCAAACAAGCTCGTCCCATCATACTGCC 296
Oy 62 CysVa1GIuGlua1aCyProLeuaAspa1aRgl1eserVa1SerAenIIeVa1Va1Asp 81
|||::: ||||| ||||| :|||
Db 297 TGCTTAGACCGAAGCTGCCAGGTTGATGACTACTCAAGGCCCAAGAAGCTGCCGCGAC 356
Oy 82 GInCyS-----SerLySa1aGlyVa1Pro1IaSp 91
||| ||||| |||
Db 357 GCCTGCAAGCCTTACTGCTGCCACCGCTCGGCTCCGCTCCCATCCACGAG 416
Oy 92 -----IleProVa1AspThrThra1a1aProGIuProSeRgIUthr 106
|||::: ||||| :|||
Db 417 AGTCGCGCTGTGTTACCGTCAGCATGACACTCTTATCACTGGCTCCGATGCTTCCC 476
Oy 107 A1aGIuPProThra1aGIuPProThrgIuGIuGIuPProThra1aGIuPProThra1aGIu-ProtH 126
||| :||| :||| :|||
Db 477 GCCATCTTCTCAACATCTCCGTGATGAAAGCCATCCCTCTGTTACTCCCGCTCCGCG 536
Oy 126 rAlaGIuPProThriaGIuPro-----ThrgIuGIuPProThra1aVa 140
||| ||||| |||
Db 537 AACGMAcCa1GaCCAAAGAcCaTaCTaCGAGAGCTGaTCCGaAGAGCG-CGTrGCAC 595
Oy 140 lProThrgIyThrgIyGIyGIyVa1ProThrgIyThGlySerPherThVa1ThrgIyar 160
||| :||| :||| :|||
Db 596 ACCCACGGGAACCTCTGGCGCG-----GGCGGGCGGCGGCAcCaCTGGCGGCTGC 646
Oy 160 gProThrlaSerThrProAlaGIuPneProGIAlaGlySerAenVa1ArGa1aserva 180
||| :||| ||||| :|||
Db 647 TAGCAGTCTCCCTACCGACGCT-----GGTGGTGAThCrCTGGCGCTGTCCCAT 697
Oy 180 lGIyGIy1IleA1aa1a1aLeuLeuGIyLeuA1a 191
||| :||| ||||| :|||
Db 698 GGCTCTCTCCCATTTATTGCTGCGCATGCT 731

RESULT 16
BM872390 688 bp mRNA linear EST 06-MAY-2003
LOCUS mgmY002xP06.f.b Magnaporthe grisea MY Uni-Zap XR library Magnaporthe
DEFINITION grisea cDNA clone mgmY002xP06 5', mRNA sequence.
ACCESSION BM872390
VERSION BM872390.2 GI:30406827
KEYWORDS EST.

ORGANISM Magnaporthe grisea (anamorph: Pyricularia grisea)
SOURCE Magnaporthe grisea
Isolate/Ectoparasite: Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes; incertae sedis; Magnaportheaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 688)
AUTHORS Ebbole,D.J., Yuan,J., Thomas,T.L., Bobrowsicz,P., Lu,G., Bhatteai
,K., and Dean,R.A.

TITLE Expressed sequence tags from the rice blast fungus, Magnaporthe
grisea

JOURNAL Unpublished
COMMENT On Mar 7, 2002 this sequence version replaced gi:19240072.
Contact: Ebbole DJ
Department of Plant Pathology & Microbiology
Texas A&M University
Peterson Bldg, MS2332, College Station, TX 77843-2132, USA
Tel: 979 845 4831

Score: 176.00 Matches: 60
 Percent Similarity: 40.39% Conservative: 22
 Best Local Similarity: 29.56% Mismatches: 91
 Query Match: 17.20% Indels: 30
 DB: 12 Gaps: 10

US-10-081-935-4 (1-194) x BM872513 (1-701)

```

Qy 7 Leu1leAlaLeuValAlaIaGlyLeuValaSerAlaGln---LeuProAsp1leProPro 25
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 5 CTTCTGTGCTCAGTTCAGACAGCAGTCTGCGCTTGCCAGACATCAACCTGTCCCGGCA 64
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 26 CysAlaLeuAsnGlyCysPheValaGluAlaLeuGlyAsnAspGlyCysThrArgLeuThrAsp 45
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 65 TGGCGGCAAGGCTTCTTCAACAGCATGAAGGACATTTGGTGTGCTCT---GCCGAC 121
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 46 PheGlyCysHisCysSerGlyProGluLeuProGlyGlnIleThrProCysValaGluGlu 65
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 122 CAGAAAGTCCCTGTGCGAAGAACCCGACTATCGCAACGGCTCGAAGACTGTCTGACTCA 181
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 66 AlaCysProLeuAspAlaArgIle-----SerValSerAsnIleValaValaSpGln--- 82
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 182 AGCTGTGCGCTGATGCAACAGAGTCCACCTGCTGAAACACATACGTGATGCTGTC 241
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 83 CysSerGlyAspGlyValProIleAspIleProProValaAspThrThrAlaAlaProGlu 102
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 242 TGGCTGTGCTGCTGCTGCT---CCGCCAGCTGCTGCCACATCTACTCTTCCA 292
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 103 ProSerGluThrAlaGluProThrAlaGluProThrGluGluProThrAlaGluProThr 122
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 293 GCCTGTGCACT-----CGAGCATGCACTCGGCTCATTTGACTCGGCGGCAACAC 346
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 123 Ala-----GluProThrAlaGluProThrHisGluProThrGluGluPro 137
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 347 TCAGTCTGCTGATCTACTACCTGCTGCTGCTGACACACACCTAGCTCAACGACGCTG 406
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 138 ThrAlaValProThr-----GlyThrGlyGlyGlyValaProThrGlyThrGlySer 154
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 407 ACTGCTGCAACGACTTCTACTAGCAGCAGCTGCTGCTGCTGCAACAGCAGCAGCAGC 466
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 155 PheThrValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyValaGly--- 173
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 467 AGCACTACTCTCTCGGCGACCTCGTCAGCAACTGCTGCG-----CCGCGCTCAGTGAG 520
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 174 -----SerAsnValaArgAlaSerValaGlyGlyLea1a 184
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 521 AAATCAGCTGAGGGGCTATCGAACAACAGCCAAATCGGTATTGCTGTGGCGCGCTGTG 580
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Qy 185 AlaAlaLeu 187
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 581 GCGGAACCTT 589
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

RESULT 18
 CD035585 640 bp mRNA linear EST 07-MAY-2003
 LOCUS mgmt020xex10.f.b Mated culture Magnaporthe grisea cDNA clone
 DEFINITION mgmt020xex10 5', mRNA sequence.
 ACCESSION CD035585
 VERSION CD035585.1 GI:30417423
 KEYWORDS EST.
 SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)
 ORGANISM Magnaporthe grisea
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes; incertae sedis; Magnaportheaceae; Magnaporthe.
 1 (bases 1 to 640)
 Ebbole,D.J., Yuan,J., Thomas,T.L., Bobrowicz,P., Lu,G., Bhatterai
 ,K. and Dean,R.A.
 TITLE Expressed sequence tags from the rice blast fungus, Magnaporthe
 grisea
 JOURNAL Unpublished
 COMMENT Contact: Ebbole DJ
 Department of Plant Pathology & Microbiology
 Texas A&M University

FEATURES

source
 location/Qualifiers
 1..640
 /organism="Magnaporthe grisea"
 /mol_type="mRNA"
 /strain="4091-5-8 X 4136-4-3"
 /db_xref="taxon:148305"
 /clone="mgmt020xex10"
 /sex="Mat1-2 and Mat1-1 mixed culture"
 /cell_type="mixed sexual development"
 /dev_stage="asci, ascospores, perithecia, mycelium"
 /clone_vector="pBluescriptSK; Site 1: EcoRI; Site 2: XhoI;
 Two mating types were co-cultivated over a filter paper on
 oatmeal agar medium. After three days at 25 C plates were
 transferred to 21 C. Perithecia with asci and ascospores
 formed at the beginning of the third week. Material was
 collected by scraping tissue from the filter paper.
 Sequences were processed by one of two methods. Where a
 full-length alignment to the M. grisea genome sequence was
 available, the EST sequence was trimmed according to the
 alignment, otherwise sequence quality was assessed using
 phredphrap version 991019 and trimmed according to phd
 files (0.05) and for vector seqs."

BASE COUNT 129 a 212 c 177 g 122 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.000477 Length: 640
 Score: 172.50 Matches: 58
 Percent Similarity: 43.46% Conservative: 25
 Best Local Similarity: 30.37% Mismatches: 67
 Query Match: 16.86% Indels: 41
 DB: 14 Gaps: 11
 US-10-081-935-4 (1-194) x CD035585 (1-640)
 Qy 1 MetGlnPheSerHisAlaLeu1e-----AlaLeuValAlaIaGlyLeuAlaSer 17
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Db 54 ATGAAGTTCAACAGCTGCTCTCTCCGCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 113
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Qy 18 AlaGln-----LeuProAsp1leProProCysAlaLeuAsnGlyPhe 31
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Db 114 GCCGAGAGCTGCATCAGCGTCCGCGCTCGGCCATCCGCTGCGCGCAGCCGCTTTC 173
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Qy 32 ValGluAlaLeuGlyAsnAspGlyCysThrArgLeuThrAspPheGlyCysHisGly--- 50
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Db 174 CTCAACGCGCGCCGACCATTTGCTGCTCGCGC---ACCGACTTCAGATGCGCAGTGCAG 230
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Qy 51 SerGlyPProGluLeuProGlyGlnIleThrProCysValaGluGluAlaCysPro----- 68
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Db 221 CAGCAGGCGCAAGATGTTTCCGCCCTCGAGAGCTCGTCAAAAGAGCTGCCAGAGCT 290
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Qy 69 ---LeuAspAlaArgIleSerValSerAsnIleValaValaAspGlnCysSerGlyAlaGly 87
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Db 291 GAGTTCCAAAGACCATTTGACGATCTGACAAGGCTCTC---ACGTGCGCTTCTGAGAGT 347
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Qy 88 ValProIleAsp1leProProValaAspThrThrAla1a----- 100
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Db 348 -----CCGCTTCAACAACGCGCGGCGCGCGGAAATACCTCAAC 389
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
 Qy 101 -----ProGluProSerGluThrAlaGluProThr-----AlaGlu 112
 |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

Db      390 CCGTCCAGCTTCATCCAGAGCCCGGAGCTGACTGCTTCCAGCACCAGCGTGCAGCC 449
Oy      113 ProthrgluGluProthralaGluProthralaGluProthralaGluProthriaglu 132
Db      450 CCGAGCGGAGACACCTCTGCGCGAGCTCTGCGGGGCGCCAGCGCCGCAAAACATGAGCC 509
Oy      133 ProthrgluGluProthralaValProthrglyGlyGlyValPro----- 149
Db      510 GCTGCGAG-----TGACAGCATCTGGTGTGTGCTGCCGCTGTGTGTGCTTGGGTAGCC 563
Oy      150 ThrglyThrglySerPheThraValThrglyArg 160
Db      564 CTGGGCTTGTGAGCATGTGACAAACTCGGGCGCA 596

RESULT 19
BU383609/c BU383609 612 bp mRNA linear EST 08-MAR-2002
LOCUS      BU383609 Dictyostelium discoideum cDNA library, CF Dictyostelium
DEFINITION discoideum cDNA clone ddc5j18 3', mRNA sequence.
ACCESSION  BU383609
VERSION    BU383609.1 GI:19292993
KEYWORDS   EST.
SOURCE     Dictyostelium discoideum
ORGANISM   Dictyostelium discoideum
REFERENCE  Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
AUTHORS    Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE      Full length cDNA of Dictyostelium discoideum at the culmination
          stage
JOURNAL    Unpublished
COMMENT    Contact: Tadasu Shin-i
          Center For Genetic Resource Information
          National Institute of Genetics
          1111 Yatae, Mishima, Shizuoka 411-8540, Japan
          Tel: 81-559-81-6856
          Fax: 81-559-81-6855
          Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
            source          1..612
                           /organism="Dictyostelium discoideum"
                           /mol_type="mRNA"
                           /strain="AX4"
                           /db_xref="taxon:44689"
                           /clone="ddc5j18"
                           /sex="mat A"
                           /dev_stage="Culmination stage"
                           /clone_id="Dictyostelium discoideum cDNA library, CF"
BASE COUNT 142 a 90 c 131 g 243 t 6 others
ORIGIN
Alignment Scores:
Pred. No.: 0.000926 Length: 612
Score: 168.50 Matches: 40
Percent Similarity: 48.96% Conservative: 7
Best Local Similarity: 41.67% Mismatches: 47
Query Match: 16.47% Indels: 2
DB: 12 Gaps: 1

US-10-081-935-4 (1-194) x BU383609 (1-612)

Oy      89 ProtleasplleProProValAspThrThraAlaAlaProGluProSerGluThraAlaGlu 108
Db      602 CCAACAGAGACTCCAACTGAGACTCAAGACTCAAGACTCAAGACTCAAGACTCAAGACT 543
Oy      109 ProthralaGlu-ProthrgluGluProthralaGluProthralaGluProthralaGlu 128
Db      542 CCAACAGAACTCCAACTGAGACTCCAACTGAGACTCCAACTGAGACTCCAACTGAGAAAC 483
Oy      128 uProThriagluProthrgluGluProthralaValProthrglyThrglyGlyVal 148
Db      482 TCCAACGAAACTCCAACTGAGAACTCCAACTGAGAACTCCAACTGAGAACTCCAACTGAGAAAC 423

```

```

Oy      148 lProthrglyThrglySerPheThraValThrglyArgProthralaSerThraAlaGlu 168
Db      422 TCCAACGAGACTCCAACTGAGAACTTACTTCAACCCCAACAGTA---ACACCAACTGA 366
Oy      168 uPheProGlyAlaGlySerAanValArgAlaSerValGlyGlyle 183
Db      365 AACTCATCAAGTGAGAACTTTATCATATCATATAAAGTGATTA 320

RESULT 20
CNS05RNM
LOCUS      CNS05RNM 1038 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetradodon nigroviridis genome survey sequence T7 end of clone
          014N06 of library C from Tetradodon nigroviridis, genomic survey
          sequence.
ACCESSION  AL350744.1 GI:8244514
VERSION    AL350744.1 GI:8244514
KEYWORDS   GSS: genome survey sequence.
SOURCE     Tetradodon nigroviridis
ORGANISM   Tetradodon nigroviridis
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
          Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
          Tetraodontidae; Tetraodon.
AUTHORS    Roest Crolius, H., Jailon, O., Dasilva, C., Bouneau, L., Fisher, C.,
          Bernot, A., Fizames, C., Wincker, P., Broctier, P., Quetier, F.,
          Saurin, W. and Weissenbach, J.
TITLE      Estimate of human gene number provided by genome-wide analysis
          using Tetradodon nigroviridis DNA sequence
JOURNAL    Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE    20296633
PUBMED     10835645
REFERENCE  2
          Roest Crolius, H., Jailon, O., Dasilva, C., Ozouf-Costaz, C.,
          Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
          Saurin, W., Bernot, A. and Weissenbach, J.
          Characterization and repeat analysis of the compact genome of the
          freshwater pufferfish Tetradodon nigroviridis
          Genome Res. 10 (7), 939-949 (2000)
JOURNAL    MEDLINE
PUBMED     20359837
REFERENCE  3 (bases 1 to 1038)
          108999143
AUTHORS    GenomeScope.
TITLE      Direct Submission
JOURNAL    Submitted (12-APR-2000) GenomeScope - Centre National de Sequencage :
          BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
          - Web : www.genoscope.cns.fr)
          This sequence is a single read and was generated as part of a large
          scale clone-end sequencing project of the Tetradodon nigroviridis
          genome. For more information, please take a look at
          http://www.genoscope.cns.fr/Tetradodon.
FEATURES   Location/Qualifiers
            source          1..1038
                           /organism="Tetradodon nigroviridis"
                           /mol_type="Genomic DNA"
                           /db_xref="taxon:99883"
                           /clone="014N06"
                           /clone_id="C"
                           /note="Genoscope sequence ID : C07C014DG03C1-end : T7"
BASE COUNT 295 a 253 c 248 g 226 t 16 others
ORIGIN
Alignment Scores:
Pred. No.: 0.00181 Length: 1038
Score: 168.00 Matches: 32
Percent Similarity: 51.16% Conservative: 12
Best Local Similarity: 37.21% Mismatches: 42
Query Match: 16.42% Indels: 0
DB: 29 Gaps: 0

US-10-081-935-4 (1-194) x CNS05RNM (1-1038)

Oy      86 AlaGlyAlaProtleasplleProProValAspThrThraAlaAlaProGluProSerGlu 105

```



```

Db      561 GGTCTCTGCGCCCAAGGCTCTCGTCC-----CTTCTCATGGCTGCTTTC 608
RESULT 24
LOCUS   BU065834                      634 bp    mRNA    linear    EST 26-AUG-2002
DEFINITION Fgr_8_K04_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA
ACCESSION BU065834
VERSION   BU065834.1  GI:22506123
KEYWORDS  EST.
SOURCE    Gibberella zeae
ORGANISM  Gibberella zeae
          Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
          Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE 1 (bases 1 to 634)
AUTHORS   Trail,F., Xu,J.-R., San Miguel,P., Halgen,R.G. and Kistler,H.C.
TITLE     Analysis of expressed sequence tags from Gibberella zeae (anamorph
          Fusarium graminearum)
JOURNAL   Fungal Genet. Biol. 38 (2), 187-197 (2003)
MEDLINE   22508120
PUBMED    12620255
COMMENT   Contact: Frances Trail
          Department of Plant Biology
          Michigan State University
          East Lansing, MI 48824, USA
          Tel: 517 432 2939
          Fax: 517 353 1926
          Email: trail@msu.edu
          Plate: 8 row: K column: 04.
          Location/Qualifiers
            1. 634
              /organism="Gibberella zeae"
              /mol_type="mRNA"
              /strain="NRRL 31084"
              /db_xref="taxon:5518"
              /clone_lib="Nitrogen-starved mycelia"
              /note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 128 a 231 c 154 g 121 t
ORIGIN
Alignment Scores:
Pred. No.: 0.00126 Length: 634
Score: 167.00 Matches: 53
Percent Similarity: 42.93% Conservative: 32
Best Local Similarity: 26.77% Mismatches: 89
Query Match: 16.32% Indels: 24
DB: 13 Gaps: 8
US-10-081-935-4 (1-194) x BU065834 (1-634)
Oy      1 MetGlnPheSerHisAlaLeuIleAlaValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      27 ATGAGTACTCGGCTGCTTGCCTCTCGCGCGCTGTCGCGCTCAGGCTAGTCTCTT 86
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Oy      21 ProAspIleProProCysAlaLeuAsnCysPheValGluAlaLeuGlyAAsnAspGlyCys 40
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      87 GCCAGCGTCCCAAGTGGCCATTCCTTGCCTCGACAAAGCCATGCCAGCAGACAGC 146
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Oy      41 ThrArgLeuThrAspPheLysCysHisCysSerLys---ProGluLeuProGlyGlnIle 59
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      147 TGTGACAGACCGACTCGCGCTGTCGTCGCAAGGGCTTCAGCGCGTCCGATCACAAGGT 206
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Oy      60 ThrProCysValGluGlnAlaCysProLeuAsnAlaAlaGlyIleSerValSerAsnIleVal 79
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      207 ACCCTTGTGTCATTGACGAGTGTGATCCGAC-----GTTGCCATCAACGAGTGC-- 257
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Oy      80 ValAspGlnCysSerLysAlaGlyValProIleAspIleProProValAspThrThrAla 99
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      258 -----CTTCCCGCTACCGAGAACTCTCC 281
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Oy      100 AlaProGluProSerGluThr--AlaGluProThrAlaGluProThrGluGluProThr 118
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

Db      282 AAGAACCCCAAGAGTGTAGAGCCCAAGTCCACCCGCGAG---GAGAGAACCCACC 338
Oy      119 AlaGluProThrAlaGluProThr-----AlaGluProThrHisGluProThrGlu 135
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      339 ACCACCGCGCGCTGCACCTTCCTGCTGCTTGTACACCAAGCGCGAGGCTTTGAG 398
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Oy      136 GluProThrAlaValProThrGlyThrGlyGlyValProThrGlyThrGlySerPhe 155
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      399 ACCACCGCGCTGCACCAACCACTGCTCCCTCCATCTCCACCAACCGCTCTCAGAG 458
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Oy      156 ThrValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyAlaGlySerAsn 175
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      459 CCGGACACCAACACCCCGCGCGCTGCACCCCTACCAAGGATCCC---GAGCAGCCAAC 515
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Oy      176 ValArgAlaSerValGlyGlyIleAlaAlaLeuLeuGlyLeuAlaAlaIle 193
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      516 GGTCTCTGCGCCCAAGGCTCTCGTCC-----CTTCCCATGGCTGCTTTC 563
RESULT 25
LOCUS   BU064409                      657 bp    mRNA    linear    EST 26-AUG-2002
DEFINITION Fgr_4_M16_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA
ACCESSION BU064409
VERSION   BU064409
KEYWORDS  EST.
SOURCE    Gibberella zeae
ORGANISM  Gibberella zeae
          Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
          Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE 1 (bases 1 to 657)
AUTHORS   Trail,F., Xu,J.-R., San Miguel,P., Halgen,R.G. and Kistler,H.C.
TITLE     Analysis of expressed sequence tags from Gibberella zeae (anamorph
          Fusarium graminearum)
JOURNAL   Fungal Genet. Biol. 38 (2), 187-197 (2003)
MEDLINE   22508120
PUBMED    12620255
COMMENT   Contact: Frances Trail
          Department of Plant Biology
          Michigan State University
          East Lansing, MI 48824, USA
          Tel: 517 432 2939
          Fax: 517 353 1926
          Email: trail@msu.edu
          Plate: 4 row: M column: 16.
          Location/Qualifiers
            1. 657
              /organism="Gibberella zeae"
              /mol_type="mRNA"
              /strain="NRRL 31084"
              /db_xref="taxon:5518"
              /clone_lib="Nitrogen-starved mycelia"
              /note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 137 a 239 c 155 g 126 t
ORIGIN
Alignment Scores:
Pred. No.: 0.00131 Length: 657
Score: 167.00 Matches: 53
Percent Similarity: 42.93% Conservative: 32
Best Local Similarity: 26.77% Mismatches: 89
Query Match: 16.32% Indels: 24
DB: 13 Gaps: 8
US-10-081-935-4 (1-194) x BU064409 (1-657)
Oy      1 MetGlnPheSerHisAlaLeuIleAlaValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      72 ATGAGTACTCGGCTGCTTGCCTCTCGCGCGCTGTCGCGCTCAGGCTAGTCTCTT 131
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Oy      21 ProAspIleProProCysAlaLeuAsnCysPheValGluAlaLeuGlyAAsnAspGlyCys 40
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      132 GCCAGCGTCCCAAGTGGCCATTCCTTGCCTCGACAAAGCCATGCCAGCAGACAGC 191
          |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```



```

Oy      41 ThrATGLeuThrAspPheLysCysHisCysSerLys---ProGluLeuProIleGlnIle 59
           |||||
Db      132 TGTGACAAAGCCGACCTCGCTGTGTCTGTCCAAGGGCTTCAGCGCTTCGATCCAAGGCT 251
           |||||
Oy      60 ThrProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleVal 79
           |||||
Db      252 ACCCTGTGTGTATGATGACGAGTGTGTGATCCGAC-----GTTCGCATCAACGAGTCC--- 302
           |||||
Oy      80 ValAaPrgLincSerLysLysAlaLysAlaProIleAspIleProProValAspThrAla 99
           |||||
Db      303 -----CTTCCCGCTAACCGAAGAACTCTGC 326
           |||||
Oy      100 AlaProGluProSerGluThr---AlaGluProThrAlaGluProThGlnGluProThr 118
           |||||
Db      327 AAGAAACCCCCCAAGAGTCTGAGGCGAAGTCCACCCCGGAG---GAGAGAAAGCCAC 383
           |||||
Oy      119 AlaGluProThrAlaGluProThr-----AlaGluProThrHisGluProThGlu 135
           |||||
Db      384 ACCACCGCGCGTCCACCTCCACTGTGTGTGTGTGCACCAACGACGCGGAGGATTTGAG 443
           |||||
Oy      136 GluProThrAlaValProThGlyThrGlyGlyValProThGlyIleThryLysPhe 155
           |||||
Db      444 ACCACCGCGCGTCCACCAACCACTGTGTCTCCCATTCATTTCCACACCGCGTCTGAGAG 503
           |||||
Oy      156 ThrValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyAlaGlySerAsn 175
           |||||
Db      504 CCGCGCACACGAGACCCCGCGCTGCTCAACCTTCACAGAGGCTCC---GACGAGCCAAAC 560
           |||||
Oy      176 ValAaGlaSerValGlyLysIleAlaAlaLeuLeuGlyLeuAlaAlaIleArg 193
           |||||
Db      561 GGTGCTGTGGCTTCAAGGCTTGTGGTGTGC-----CTTGCCATCACTGGCTTTC 608
           |||||

```

RESULT	26
B0060530	
LOCUS	B0060530
DEFINITION	B0060530 664 bp mRNA linear EST 26-AUG-2002 Fgr-C_1_E06_T3 Carbon-starved mycelia Gibberella zeae cDNA, mRNA sequence.
ACCESSION	B0060530
VERSION	B0060530.1 GI:22500819
KEYWORDS	EST.
SOURCE	Gibberella zeae
ORGANISM	Gibberella zeae
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
AUTHORS	Trail, F., Xu, J.-R., San Miguel, P., Halgren, R.G. and Kistler, H.C.
TITLE	Analysis of expressed sequence tags from Gibberella zeae (anamorph Fusarium graminearum)
JOURNAL	Fungal Genet. Biol. 38 (2), 187-197 (2003)
MEDLINE	22508120
PUBMED	12620255
COMMENT	Contact: Frances Trail Department of Plant Biology Michigan State University East Lansing, MI 48824, USA Tel.: 517 432 2939 Fax: 517 353 1926 Email: trail@msu.edu
FEATURES	Plate: 1 row: E column: 06. Location/Qualifiers
source	1. 664 /organism="Gibberella zeae" /mol_type="mRNA" /strain="NRRL 31084" /db_xref="taxon:5518" /clone_id="Carbon-starved mycelia" /note="Vector: Uni-ZapII; Site_1: EcoRI, Site_2: XhoI"
BASE COUNT	139 a 241 c 155 g 129 t
ORIGIN	
Alignment Scores:	0.00132 length: 664 red. No.:

Score:	167.00	Matches:	53
Percent Similarity:	42.93%	Conservative:	32
Best local Similarity:	26.77%	Mismatches:	89
Query Match:	16.32%	Indels:	8
DB:	13	Gaps:	24

```

Oy 1 MetGlnPheSerH1a1aLeuH1a1aLeuVala1a1aGly1aLeuAlaSer1aGln1aLeu 20
Db 80 ATGAAGTACTCGTCGCTTGTGTTGCTCTCGCGCTGTGGCCGTACAGCTACTGTCCTT 139
Oy 21 ProAsp1aLeuProCysa1aLeuAsnCysPheValaGua1aLeuG1aYasnAsp1aCys 40
Db 140 GCCAGCGTCCCAAGTAGTGCGCATTCCTTGCTCCAGCAAGGCCATGCCAGCAAGCAC 199
Oy 41 ThrArgLeuThrAspPheLysCysHisCysSerLys---ProGluLeuProGluGln1a 59
Db 200 TGTGACAAAGACCGACCTCGCGCTGTGTGTGCAAGGGCTTCACGCGTGTCCGATCCAA 259
Oy 60 ThrProCysValaGluGua1aCysProLeuAsp1aAsn1aG1aSerValaSerAsn1aLeu 79
Db 260 ACCTCTGTGTGTCATTGACAGAGTGAGTACCAC-----GTGGCCATCAACCAAGCTC 310
Oy 80 ValAspGlnCysSerLysa1aGlyValaPro1aAsp1aProCysValaAspThrThrAla 99
Db 311 -----CTTCCCGCTACCGAAGAACTTCC 334
Oy 100 AlaProGluProSerGluThr--AlaGluProThrAlaGluProThrGluGluProThr 118
Db 335 AAGAACCCTCCCAAGAGATGTGAGCGCCAGATCCACCGCCGAG--GAGAGAGAACCCACC 391
Oy 119 AlaGluProThrAlaGluProThr-----AlaGluProThrHisGluProThrGlu 135
Db 392 ACCACGCGCGCTGCACCTCCACCTTGCTGCTGTACACCAAGCAGCGGAGGTTGTGAG 451
Oy 136 GluProThrAlaValProThrGlyThrG1aGlyValaProThrGlyThrGlySerPhe 155
Db 452 ACCACGCGCGTGCACCAACACATGTGGTCCCATTCATCCACACCGCTGTAGAGG 511
Oy 156 ThrValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyAlaGlySerAsn 175
Db 512 CCGCCACCAAGACACCCCGCGCGTGCACACCTTACCAAGGGTCCC--GACACAGCCCAAC 568
Oy 176 ValArgAlaSerValaGlyGlyLea1aAla1aLeuLeuGlyLeuAlaAla1aTy 193
Db 569 GGTGCTCTGCGCTCAAGGATCTCGGTGCC-----CTCGCCATAGGCTGCTTTC 616

```

BU065008	RESULT 27
LOCUS	BU065008
DEFINITION	696 bp mRNA linear EST 26-AUG-2002
ACCESSION	For_6_G07_T3 Nitrogen-starved mycelia <i>Gibberella zeae</i> cDNA, mRNA sequence.
VERSION	BU065008
KEYWORDS	BU065008.1 GI:22505297
SOURCE	EST.
ORGANISM	<i>Gibberella zeae</i>
REFERENCE	<i>Gibberella zeae</i>
AUTHORS	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocerales; Nectriaceae; <i>Gibberella</i> .
TITLE	1 (bases 1 to 696)
JOURNAL	Trail, F., Xu, J., -R., San Miguel, P., Halgren, R. G. and Kistler, H. C.
MEDLINE	Analysis of expressed sequence tags from <i>Gibberella zeae</i> (anamorph <i>Fusarium graminearum</i>)
PUBMED	Fungal Genet. Biol. 38 (2), 187-197 (2003)
COMMENT	22508120
	12620255
	Contact: Frances Trail
	Department of Plant Biology
	Michigan State University
	East Lansing, MI 48824, USA
	Tel: 517 432 2939
	Fax: 517 353 1926

Email: trail@msu.edu
 Plate: 6 row: G column: 07.
 Location/Qualifiers
 1. .696
 /organism="Gibberella zeae"
 /mol_type="mRNA"
 /strain="NRRL 31084"
 /db_xref="taxon:5518"
 /clone_id="Nitrogen-starved mycelia"
 /note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 151 a 248 c 162 g 135 t
 ORIGIN

Alignment Scores:

Pred. No.: 0.00139 Length: 696
 Score: 167.00 Matches: 53
 Percent Similarity: 42.93% Conservative: 32
 Best Local Similarity: 26.77% Mismatches: 89
 Query Match: 16.32% Indels: 24
 DB: 13 Gaps: 8

US-10-081-935-4 (1-194) x BU065008 (1-696)

Oy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
 Db 80 ATGAGTACTCTCGTCTGCTTCTGCTCTGCGCGCTGTCAGGCTCAGTCTCTT 139
 Oy 21 ProAspIleProProCysAlaLeuAnsCysPheValGluAlaLeuGlyAsnAspGlyCys 40
 Db 140 GCCGACGTCCCAAGTGGCCCATCTCTGCTCGACAGAGGCATCGCCAGCAGACAGC 199
 Oy 41 ThrArgLeuThrAspPheLeuScyHisCysSerLys---ProGluLeuProGlyGlnIle 59
 Db 200 TGTGACAAAGACCGACCTCGCTGTGTCTGCAAGGCGCTTCAGCGCTCGATCCAAAGCT 259
 Oy 60 ThrProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAnIleVal 79
 Db 260 ACCTCTGTGTATTGACGAGTGTGTACCGAC-----GTGCCATCAACAGGTC--- 310
 Oy 80 ValAspGlnCysSerLysAlaGlyValProIleAspIleProProValAspThrThrAla 99
 Db 311 -----CTTCCCGCTAACGAGAACCTCTGC 334
 Oy 100 AlaProGluProSerGluThr---AlaGluProThrAlaGluProThrgluGluProThr 118
 Db 335 AAGAACCCGCCCAAGAGCTGAGGCCAAGTCCACCGCCGAG--GAGGAGAAGCCCAAC 391
 Oy 119 AlaGluProThrAlaGluProThr-----AlaGluProThrHisGluProThrglu 135
 Db 392 ACCACCGCGCGTCCACCTCCACCTTGTGCTGTTCACACACGCGCCGAGGTGTGAG 451
 Oy 136 GluProThrAlaValProThrglyThrGlyGlyValProThrglyThrGlySerPhe 155
 Db 452 ACCACCGCGCGTCCACCTCCACCTGCTCCATCATTCACACACCGCTCTGAGAG 511
 Oy 156 ThrValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyAlaIleSerAn 175
 Db 512 CCGCGCACGAGACCCCGCGCGTCCACCTTACCAAGGATCCC--GAGCAGAGCCAAAC 568
 Oy 176 ValArgAlaSerValGlyGlyIleAlaAlaLeuLeuGlyLeuAlaIleTyr 193
 Db 569 GGTGCTGCTGGCTCAAGGCTCTCGTGCC-----CTGCCATGCTCTTCTT 616

RESULT 28
 BU066113 699 bp mRNA linear EST 26-AUG-2002
 LOCUS Fgr 9 G08_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA
 DEFINITION
 accession
 BU066113
 VERSION BU066113.1 GI:22506402
 KEYWORDS
 EST.
 SOURCE Gibberella zeae
 ORGANISM Gibberella zeae

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

REFERENCE
 1 (bases 1 to 699)
 Trail, F., Xu, J., R., San Miguel, P., Halgren, R.G. and Kistler, H.C.
 Analysis of expressed sequence tags from *Gibberella zeae* (anamorph
Fusarium graminearum)
 Fungal Genet. Biol. 38 (2), 187-197 (2003)

JOURNAL
 MEDLINE
 PUBMED
 22508120
 12620255

COMMENT
 Contact: Frances Trail
 Department of Plant Biology
 Michigan State University
 East Lansing, MI 48824, USA
 Tel: 517 432 2939
 Fax: 517 353 1926
 Email: trail@msu.edu
 Plate: 9 row: G column: 08.
 Location/Qualifiers

FEATURES
 source
 1. .699
 /organism="Gibberella zeae"
 /mol_type="mRNA"
 /strain="NRRL 31084"
 /db_xref="taxon:5518"
 /clone_id="Nitrogen-starved mycelia"
 /note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 152 a 249 c 162 g 136 t
 ORIGIN

Alignment Scores:

Pred. No.: 0.0014 Length: 699
 Score: 167.00 Matches: 53
 Percent Similarity: 42.93% Conservative: 32
 Best Local Similarity: 26.77% Mismatches: 89
 Query Match: 16.32% Indels: 24
 DB: 13 Gaps: 8

US-10-081-935-4 (1-194) x BU066113 (1-699)

Oy 1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
 Db 82 ATGAGTACTCTCGTCTGCTTCTGCTCTGCGCGCTGTCAGGCTCAGTCTCTT 141
 Oy 21 ProAspIleProProCysAlaLeuAnsCysPheValGluAlaLeuGlyAsnAspGlyCys 40
 Db 142 GCCGACGTCCCAAGTGGCCCATCTCTGCTCGACAGAGGC/TGCCAGCAGACAGC 201
 Oy 41 ThrArgLeuThrAspPheLeuScyHisCysSerLys---ProGluLeuProGlyGlnIle 59
 Db 202 TGTGACAAAGACCGACCTCGCTGTGTCTGCAAGGCGCTTCAGCGCTCGATCCAAAGCT 261
 Oy 60 ThrProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAnIleVal 79
 Db 262 ACCTCTGTGTATTGACGAGTGTGTACCGAC-----GTGCCATCAACAGGTC--- 312
 Oy 80 ValAspGlnCysSerLysAlaGlyValProIleAspIleProProValAspThrThrAla 99
 Db 313 -----CTTCCCGCTAACGAGAACCTCTGC 336
 Oy 100 AlaProGluProSerGluThr---AlaGluProThrAlaGluProThrgluGluProThr 118
 Db 337 AAGAACCCGCCCAAGAGCTGAGGCCAAGTCCACCGCCGAG--GAGGAGAAGCCCAAC 393
 Oy 119 AlaGluProThrAlaGluProThr-----AlaGluProThrHisGluProThrglu 135
 Db 394 ACCACCGCGCGTCCACCTCCACCTTGTGCTGTTCACACACGCGCCGAGGTGTGAG 453
 Oy 136 GluProThrAlaValProThrglyThrGlyGlyValProThrglyThrGlySerPhe 155
 Db 454 ACCACCGCGCGTCCACCACTGCTGCTCCATCATTCACACACCGCTCTGAGAG 513
 Oy 156 ThrValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyAlaIleSerAn 175
 Db 514 CCGCGCACGAGACCCCGCGCGTCCACCTTACCAAGGATCCC--GAGCAGAGCCAAAC 570

```

Qy      176 ValArgAlaSerValGlyGlyIleAlaAlaLeuLeuGlyLeuAlaAlaTyr 193
Db      571 GGTGCTCTGCGCTCAAGGCTCTGTCGCC-----CTCCGCATGCGTCTTTC 618

RESULT 29
LOCUS   BU064737              718 bp    mRNA    linear    EST 26-AUG-2002
DEFINITION fgr_5_K14_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA
ACCESSION BU064737
VERSION   BU064737.1    GI:22505026
KEYWORDS  EST.
SOURCE    Gibberella zeae
ORGANISM  Gibberella zeae
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
AUTHORS   Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
TITLE      Trail,F., Xu,J.-R., San Miguel,P., Halgren,R.G. and Kistler,H.C.
           Analysis of expressed sequence tags from Gibberella zeae (anamorph
           Fusarium graminearum)
JOURNAL   Fungal Genet. Biol. 38 (2), 187-197 (2003)
MEDLINE   22508120
PUBMED    12620255
COMMENT   Contact: Frances Trail
           Department of Plant Biology
           Michigan State University
           East Lansing, MI 48824, USA
           Tel: 517 432 2939
           Fax: 517 353 1926
           Email: trail@msu.edu
           Plate: 5 row: K column: 14.
FEATURES
  source          1..718
                  /organism="Gibberella zeae"
                  /mol_type="mRNA"
                  /strain="NRRL 31084"
                  /db_xref="taxon:5518"
                  /clone_lib="Nitrogen-starved mycelia"
                  /note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      157 a      253 c      168 g      140 t
ORIGIN
Alignment Scores:
Pred. No.:      0.00144      Length:      718
Score:          167.00      Matches:      53
Percent Similarity: 42.93%      Conservative: 32
Best Local Similarity: 26.77%      Mismatches: 89
Query Match:    16.32%      Indels:      24
DB:             13      Gaps:      8
US-10-081-935-4 (1-194) x BU064737 (1-718)
Qy      1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
Db      80 ATGAAGACACCGGCGCTTTCGTCGTCGCGCGCTGTCGCGCGCTCAAGCTCACTCTT 139
Qy      21 ProAspIleProProCysAlaAlaLeuAsnCysPheValAlaAlaLeuGlyAsnAspGlyCys 40
Db      140 GCCGACGTCGCCCAAGTCGCGCATTCCTGCTCCACAGAGCATCGCCAGAGACACGAC 199
Qy      41 ThrArgLeuThrAspPheLysCysHisCysSerLys---ProGluLeuProGlyGlnIle 59
Db      200 TGTGACAAAGCCACCTCGCGCTGTGTCTGCAAGGGCTTCACGCGTGTCCATCCAAAGCT 259
Qy      60 ThrProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleVal 79
Db      260 ACCTCTGTGTCAATGACGAGTGTGTAACGAC-----GTTCGCATCAACGAGGTC--- 310
Qy      80 ValAspGlnCysSerLysAlaGlyValProIleAspIleProProValAspThrThAla 99
Db      311 -----CTTCCCGCTACCGAAGACTCTGCG 334

```

```

Qy      100 AlaProGluProSerGluThr---AlaGluProThrAlaGluProThrGluGluProThr 118
Db      335 AAGAACCCCCCAAGAGTGTGAGGCCAAGTCACCGCCGAG---GACGAGAACCCCAAC 391
Qy      119 AlaGluProThrAlaGluProThr-----AlaGluProThrHisGluProThrGlu 135
Db      392 ACCACGCCCGCTGACCTCCACTTGGTGTGTGTCACACACCGCCGAGGTGTTGAG 451
Qy      136 GluProThrAlaValAlaProThrGlyThrGlyGlyValAlaProThrGlyThrGlySerPhe 155
Db      452 ACCACCCCGCTGCGACACCATGTCGCTCCATCATTCACACACCGCTGTGAGAG 511
Qy      156 ThrValThrGlyArgProThrAlaSerThrProAlaGluPheProGlyAlaGlySerAsn 175
Db      512 CCCCACACAGACACCCCGCGCTGCCACCCCTCAACAGGCTCC---GACACGACCAAC 568
Qy      176 ValArgAlaSerValGlyGlyIleAlaAlaLeuLeuGlyLeuAlaAlaTyr 193
Db      569 GGTGCTCTGCGCTCAAGGCTCTGTCGCC-----CTCCGCATGCGTCTTTC 616

RESULT 30
LOCUS   BU064923              726 bp    mRNA    linear    EST 26-AUG-2002
DEFINITION fgr_6_C13_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA
ACCESSION BU064923
VERSION   BU064923.1    GI:22505212
KEYWORDS  EST.
SOURCE    Gibberella zeae
ORGANISM  Gibberella zeae
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
AUTHORS   Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
TITLE      Trail,F., Xu,J.-R., San Miguel,P., Halgren,R.G. and Kistler,H.C.
           Analysis of expressed sequence tags from Gibberella zeae (anamorph
           Fusarium graminearum)
JOURNAL   Fungal Genet. Biol. 38 (2), 187-197 (2003)
MEDLINE   22508120
PUBMED    12620255
COMMENT   Contact: Frances Trail
           Department of Plant Biology
           Michigan State University
           East Lansing, MI 48824, USA
           Tel: 517 432 2939
           Fax: 517 353 1926
           Email: trail@msu.edu
           Plate: 6 row: C column: 13.
FEATURES
  source          1..726
                  /organism="Gibberella zeae"
                  /mol_type="mRNA"
                  /strain="NRRL 31084"
                  /db_xref="taxon:5518"
                  /clone_lib="Nitrogen-starved mycelia"
                  /note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      159 a      256 c      168 g      143 t
ORIGIN
Alignment Scores:
Pred. No.:      0.00146      Length:      726
Score:          167.00      Matches:      53
Percent Similarity: 42.93%      Conservative: 32
Best Local Similarity: 26.77%      Mismatches: 89
Query Match:    16.32%      Indels:      24
DB:             13      Gaps:      8
US-10-081-935-4 (1-194) x BU064923 (1-726)
Qy      1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
Db      88 ATGAAGACACCGGCGCTTTCGTCGTCGCGCGCTGTCGCGCGCTCAAGCTCACTCTT 147
Qy      21 ProAspIleProProCysAlaAlaLeuAsnCysPheValAlaAlaLeuGlyAsnAspGlyCys 40

```

```

Db      148 GCCGACGTCCTCCAGTGCCTGCTCTTCTGCTGACACAGGCATTCGCCACGAGACCAGC 207
      41 ThrArgLeuThrAspPheLysCysHisCysSerLys---ProGluLeuProGlyGlnIle 59
      208 TGTGACAGAGCCGACCTCGCTGTCTGTGACAGGCGCTTACAGCGCTGTCCATCCAGCGCT 267
      60 ThrProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleVal 79
      268 ACCTCTGTGTCTCATTTGACGAGTGTGTGATCCGAC-----GTTCGCATCAACGAGATC--- 318
      80 ValAspGlnCysSerLysValGlyValProIleAspIleProProValAspThrThrAla 99
      319 -----CTTCCCGCTACCGAGAACCTCTGC 342
      100 AlaProGluProSerGluThr---AlaGluProThrAlaGluProThrGluProThr 118
      343 AAGAACCCGCCCAAGAGTGTGAGGCCAGATCCACCCCGAG---GAGGAGAACCCACCC 399
      119 AlaGluProThrAlaGluProThr-----AlaGluProThrHisGluProThrGlu 135
      400 ACCACCGCCGCTGCGCACCCTCCACCTGTGCTGTGTCCACACACGCGCGAGTTGTGAG 459
      136 GluProThrAlaValProThrGlyThrGlyGlyValProThrGlyThrGlySerPhe 155
      460 ACCACCGCCGCTGCGCACCCTGCTGCTGCTCCATCTCCATCCACACCGCGCTGAGAG 519
      156 ThrValThrGlyAProThrAlaSerThrProAlaGluPheProGlyAlaGlySerAsn 175
      520 CCGCGCCACGACACCCCGCGCTGCGCACCCCTACCAAGGGTCCC---GAGCAGGCCAAC 576
      176 ValArgAlaSerValGlyGlyIleAlaAlaAlaLeuLeuGlyLeuAlaIleTyr 193
      577 GTGCTCTGCTGCGCTCAAGGGTCTCGTGCC-----CTCCGCATGCGCTCTTTC 624

RESULT 31
CD458964      841 bp      mRNA      linear      EST 03-JUN-2003
DEFINITION    Fg08_01h09_A Fg08_AAFC_ECORC_Fusarium_graminearum_complex_substrate
GIBBERELLA   zaeae cDNA clone Fg08_01h09, mRNA sequence.
ACCESSION    CD458964
VERSION      CD458964.1 GI:31373704
KEYWORDS     EST.
SOURCE       Gibberella zaeae
ORGANISM     Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
              Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE    1 (bases 1 to 841)
AUTHORS      Watson,R.J., Heyes,R., Chapados,J., Couroux,P., Harris,L.J., Hattori
              J., Lacroix,C., Ouellet,T., Robert,L.S., Singh,J.A., Sprott,D. and
              Tinker,N.A.
              A cDNA library prepared from Fusarium graminearum grown on a
              complex plant substrate
              Unpublished
              Contact: Watson, Robert.J.
              Eastern Cereal and Oilseed Research Centre
              Agriculture and Agri-food Canada
              Bldg. 20, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
              CANADA
              Tel: (613) 759-1655
              Fax: (613) 759-1701
              Email: watsonrj@agr.gc.ca.

FEATURES
  Source
    1..841
    /organism="Gibberella zaeae"
    /mol_type="mRNA"
    /strain="DAOM 180378"
    /db_xref="taxon:5518"
    /clone="Fg08_01h09"
    /tissue_type="mycelium"
    /dev_stage="asexual"
    /lab_host="E. coli DH10B"
    /clone_id="Fg08_AAFC_ECORC_Fusarium_graminearum_complex_s

```

```

BASE COUNT    191 a      270 c      200 g      179 t
ORIGIN
ubstrate"
/Note="Vector: pBluescript II+; Site 1: ECORI; Site 2:
XhoI; Fusarium graminearum grown on a complex plant
substrate-- wheat leaves treated to remove most of the low
molecular weight, water-soluble components."
Alignment Scores:
Pred. No.:      0.00172      Length:      841
Score:          167.00      Matches:      53
Percent Similarity: 42.93%      Conservative: 32
Best Local Similarity: 26.77%      Mismatches: 89
Query Match:    16.32%      Indels:      24
DB:              14      Gaps:      8

US-10-081-935-4 (1-194) x CD458964 (1-841)
      1 MetGlnPheSerHisAlaLeuIleAlaLeuValAlaIleGlyLeuAlaSerAlaGlnLeu 20
      82 ATGAGTACTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 141
      21 ProAspIleProProCysAlaLeuAsnCysPheValGluAlaLeuGlyAsnAspGlyCys 40
      142 GCCGAGTCCGCCAAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 201
      41 ThrArgLeuThrAspPheLysCysHisCysSerLys---ProGluLeuProGlyGlnIle 59
      202 TGTGACAGAGCCGACCTCGCTGTGTGTGCAAGGCTTACAGCGCTGTCCATCCAGCGCT 261
      60 ThrProCysValGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleVal 79
      262 ACCTCTGTGTCTCATTTGACGAGTGTGTGATCCGAC-----GTTCGCATCAACGAGATC--- 312
      80 ValAspGlnCysSerLysValGlyValProIleAspIleProProValAspThrThrAla 99
      313 -----CTTCCCGCTACCGAGAACCTCTGC 336
      100 AlaProGluProSerGluThr---AlaGluProThrAlaGluProThrGluProThr 118
      337 AAGAACCCGCCCAAGAGTGTGAGGCCAGATCCACCCCGAG---GAGGAGAACCCACCC 393
      119 AlaGluProThrAlaGluProThr-----AlaGluProThrHisGluProThrGlu 135
      394 ACCACCGCCGCTGCGCACCCTGCTGCTGCTCCATCTCCATCCACACCGCGCTGAGAG 453
      136 GluProThrAlaValProThrGlyThrGlyGlyValProThrGlyThrGlySerPhe 155
      454 ACCACCGCGCTGCGCACCCTGCTGCTGCTCCATCTCCATCCACACCGCGCTGAGAG 513
      156 ThrValThrGlyAProThrAlaSerThrProAlaGluPheProGlyAlaGlySerAsn 175
      514 CCGCGCCACGACACCCCGCGCTGCGCACCCCTACCAAGGGTCCC---GAGCAGGCCAAC 570
      176 ValArgAlaSerValGlyGlyIleAlaAlaLeuLeuGlyLeuAlaIleTyr 193
      571 GTGCTCTGCTGCGCTCAAGGGTCTCGTGCC-----CTCCGCATGCGCTCTTTC 618

RESULT 32
CD458020      867 bp      mRNA      linear      EST 03-JUN-2003
DEFINITION    Fg08_01h09_R Fg08_AAFC_ECORC_Fusarium_graminearum_complex_substrate
GIBBERELLA   zaeae cDNA clone Fg08_01h09, mRNA sequence.
ACCESSION    CD458020
VERSION      CD458020.1 GI:31372760
KEYWORDS     EST.
SOURCE       Gibberella zaeae
ORGANISM     Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
              Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE    1 (bases 1 to 867)
AUTHORS      Watson,R.J., Heyes,R., Chapados,J., Couroux,P., Harris,L.J., Hattori
              J., Lacroix,C., Ouellet,T., Robert,L.S., Singh,J.A., Sprott,D. and

```

Tinker, N.A.	
A cDNA library prepared from <i>Fusarium graminearum</i> grown on a complex plant substrate	
Unpublished	
Contact: Watson, Robert J.	

FEATURES	Location/Qualifiers
source	1. 867

US-10-081-935-4 (1-194) x CD458020 (1-867)

Qy	1	MeGlnPheSerHisAlaLeuIleAlaLeuValAlaIleArgLysLeuAlaSerAlaGlnLeu	20
Db	82	ATGAATACTACCGTCGCTTCGTGCGCTCGCGCCGTGTTGCCGCTCAGGCTCACTTCTT	141
Qy	21	ProAspIleProProCysAlaLeuAsnCysPheValGluAlaLeuGlyValAsnAspIlyCys	40
Db	142	GCCGACGCTCCCAAGTGCAGCTTCCTCTTCGACCAAGGCATCGCAGCAGACACG	201
Qy	41	ThrArgLeuThrAspPheCysHisIAscSerLys---ProGluLeuProGlnIle	59
Db	202	TGTGACAAAGCCGACTTCGCTGCTGTCTGTCGAAGGCTTCAGCGCTGTCGATCCAAAGCT	261
Qy	60	ThrProCysValGluGluValIaCysProLeuAspAlaArgIleSerValSerAsnIleVal	79
Db	262	ACCTCTGTGTGATTATGACGAGTGTGGTACCGAC-----GTTCCGATCAACGAGATC---	312
Qy	80	ValAspGlnCysSerLysLeuIaGlyValProIleAspIleProProValAspThrThrAla	99
Db	313	-----CTTCCCGCTACCGAGAACCTCTGC	336
Qy	100	AlaProGluProSerGluThr---AlaGluProThrAlaGluProThrGlnGluProThr	118
Db	337	AGCAACCCCCCAAGAGATGTGAGGCCAAATCACCACCGCGAG--GAGGAGAACCCAC	393
Qy	119	AlaGluProThrAlaGluProThr-----AlaGluProThrHisGluProThrGlu	135
Db	394	ACCACCGCGCTGCACACTCCACTTGGTCGTTCACACACGAGCGCGAGGTTGTGAG	453
Qy	136	GluProThrAlaValProThrGluThrArgIleGlyValProThrGlnThrGlySerPhe	155
Db	454	ACCACCGCGCTGCACACCACTGTGCTCCCATCATTCACACACCGCTGTGAGAG	513

[illegible]

RESULT 33

BM868215

LOCUS
DEFINITION

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANIS

REFERENCE

AUTHORS

.....

TITLE

JOURNAL
COMMENT

COMMENT

Expressed sequence tags from the rice blast fungus, *Magnaporthe grisea*
 Unpublished
 On Mar 7, 2002 this sequence version replaced gi:19235897.
 Contact: EDDOLE DJ
 Department of Plant Pathology & Microbiology
 Texas A&M University
 Peterson Bldg, MS2132, College Station, TX 77843-2132, USA
 Tel: 979 845 4831
 Fax: 979 845 6483
 Email: d-eddole@tamu.edu
 Chromatogram file of this sequence is available, see contact person
 .Bert n hit (April. 22, 2003) gbl/AAH6312.1 | MACE interacting
 .Protein I; Acili|Magnaporthe gr.. 179.3e-44

FEATURES

SOURCE

```

Plate: mgc8013 row: P column: 15
Seq primer: T3.
Location/Qualifiers
1. .606

```

BASE COUNT
ORIGIN

115 a	200 c	171 g	120 e
-------	-------	-------	-------

Alignment Scores:

Pred. No.

Score:

Percent S
Best Loca

Query Mat.
DB:

length:	600
matches:	54
Conservative:	22
Similarity:	67
Mismatches:	34
Indels:	9
Gaps:	3
	12
	166.00
	42.9%
	30.51%
	16.2%
	0.0043

DB:

COMMENT

Contact: Watson, Robert.J.
 Eastern Cereal and Oilseed Research Centre
 Agriculture and Agri-Food Canada
 Bldg. 20, Central Experimental Farm, Ottawa, Ontario, KIA 0C6,
 CANADA
 Tel: (613) 759-1655
 Fax: (613) 759-1701
 Email: watsonrj@agr.gc.ca

FEATURES

Location/Qualifiers
 1. 822

BASE COUNT 184 a 267 c 195 g 172 t 4 others
 ORIGIN

Alignment Scores:

Pred. No.: 0.002 Length: 822
 Score: 166.00 Matches: 54
 Percent Similarity: 43.22% Conservative: 32
 Best Local Similarity: 27.14% Mismatches: 88
 Query Match: 16.23% Indels: 25
 DB: 14 Gaps: 8

US-10-081-935-4 (1-194) x CD460963 (1-822)

QY 1 MetGlnPheSerHisAlaLeu1AlaLeuValAlaGlyLeuAlaSerAlaGlnLeu 20
 DB 82 ATGAGACTCGCTCGCTTCGTCGCTCGCCGCTTCGCCGCTCAGCTCCTT 141
 QY 21 ProAsp1LeProProCysAlaLeuAsnCysPheValGlnAlaLeuGlyAsnAspGlyCys 40
 DB 142 GCCGACGTCCCGCAAGTCCGCTTCCTTCGTCGACAGGCATCGCCAGCAGACGAGC 201
 QY 41 ThrArgLeuThrAspPheLysCysHisCysSerLys---ProGluLeuProGlyGlnIle 59
 DB 202 TGTGACAAAGCCACCTCGCTGTGTGTCAGAGGCTTCAGCGCTGTCATCCATCCAGGCT 261
 QY 60 ThrProCysValGlnGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleVal 79
 DB 262 ACCTCTGTGTGTCATTGACGAGTGTGTAACGAC-----CTTCCATCAACAGAGTC--- 312
 QY 80 ValAspGlnCysSerLysAlaGlyValProIleAsp1LeProProValAlaPthrThrAla 99
 DB 313 -----CTTCCGCTACCGAAGCTCTGC 336
 QY 100 AlaProGluProSerGluThr---AlaGluProThrAlaGluProThrGluGluProThr 118
 DB 337 AAGAACCCCCCAAGAGTGTGAGGCCAAGTCCACGCCGAG---GAGGAGAGCCGAC 353
 QY 119 Ala-GluProThrAlaGluPro-----ThrAlaGluProThrHisGluProThrG1 135
 DB 394 ACCACCGCGCTGCACCTCCCTGTGTCGTCGTCACCAAGGCGCGAGTTGTGA 453
 QY 135 uGluProThrAlaValProThrGlyThrGlyGlyValProThrGlyThrGlySerPh 155
 DB 454 GACCAACCGCGCTGCACCACTGTGCTCCATCATTCACCAACCGCTCTGAGGA 513
 QY 155 eThrValThrGlyArgProThrAlaSerThrProAlaGluPhProGlyValaGlySerAs 175
 DB 514 GCCCGCACCAAGACCCCGCGCTGTCACCTTACCAAGAGTCCC---GAGCAGGCCAA 570
 QY 175 nValaGlnCysSerLysAlaGlyValProIleAsp1LeProProValAlaPthrThrAla 193

DB

571 CGGTCGCTCGCTCAGAGTCTCGCTGCC-----CTGCCATGCTGCTTTC 619

RESULT 36

BU064419

611 bp mRNA linear EST 26-AUG-2002

LOCUS

BU064419

DEFINITION

FGF_4_N02_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA

ACCESSION

BU064419.1 GI:22504708

VERSION

EST.

KEYWORDS

Gibberella zeae

SOURCE

Gibberella zeae

ORGANISM

Gibberella zeae

REFERENCE

Ekartoyca, Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

AUTHORS

Trail, F., Xu, J., R., San Miguel, P., Halgren, R. G. and Kistler, H. C.

TITLE

Analysis of expressed sequence tags from Gibberella zeae (anamorph

JOURNAL

Fungal Genet. Biol. 38 (2), 187-197 (2003)

MEDLINE

22508120

COMMENT

Contact: Frances Trail
 Department of Plant Biology
 Michigan State University
 East Lansing, MI 48824, USA
 Tel: 517 432 2939
 Fax: 517 353 1926
 Email: trail@msu.edu
 Plate: 4 row: N column: 02.

FEATURES

Location/Qualifiers
 1. 611

BASE COUNT 128 a 229 c 143 g 111 t
 ORIGIN

Alignment Scores:

Pred. No.: 0.00188 Length: 611
 Score: 164.50 Matches: 52
 Percent Similarity: 40.61% Conservative: 28
 Best Local Similarity: 26.40% Mismatches: 88
 Query Match: 16.08% Indels: 29
 DB: 13 Gaps: 7

US-10-081-935-4 (1-194) x BU064419 (1-611)

QY 1 MetGlnPheSerHisAlaLeu1AlaLeuValAlaGlyLeuAlaSerAlaGlnLeu 20
 DB 80 ATGAGACTCGCTCGCTTCGTCGCTCGCCGCTTCGCCGCTCAGCTCCTT 139
 QY 21 ProAsp1LeProProCysAlaLeuAsnCysPheValGlnAlaLeuGlyAsnAspGlyCys 40
 DB 140 GCCGACGTCCCGCAAGTCCGCTTCCTTCGTCGACAGGCATCGCCAGCAGACGAGC 199
 QY 41 ThrArgLeuThrAspPheLysCysHisCysSerLys---ProGluLeuProGlyGlnIle 59
 DB 200 TGTGACAAAGCCACCTCGCTGTGTGTCAGAGGCTTCAGCGCTGTCATCCATCCAGGCT 259
 QY 60 ThrProCysValGlnGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleVal 79
 DB 260 ACCTCTGTGTGTCATTGACGAGTGTGTAACGAC-----CTTCCATCAACAGAGTC--- 310
 QY 80 ValAspGlnCysSerLysAlaGlyValProIleAsp1LeProProValAlaPthrThrAla 99
 DB 311 -----CTTCCGCTACCGAAGCTCTGC 334
 QY 100 AlaProGluProSerGluThr---AlaGluProThrAlaGluProThrGluGluProThr 118

```

Db      335 AAGAACCCCAAGAGTCTGAGGCCAAGTCCACCCGAG--GAGGAGAACCCACC 391
Qy      119 AlagluProthralagluProthr-----AlagluProthrAlagluProthglu 135
          ||| ||| |||
Db      392 ACCAGCCCGCGCTGCACCTTCGCTGCTTGTCAACACACAGCGCGGCTGTGTAG 451
Qy      136 GluProthralaValProthrglyThrglyglYvalProthrglyThrglySerphe 155
          ||| ||| |||
Db      452 ACCACCCCGCTGCACCACTGCTGCTCCATCTCCACACCGCTGCTGAGAG 511
Qy      156 ThrValThrglyArProthralaSerThProAlagluPheProglYAlaglySerAsn 175
          ||| ||| |||
Db      512 CCGCCACACAGACACCCCGCGCTGCCACCCCTTACCAAGGCTCCGAGACAGCCACGCT 571
Qy      176 ValArgAlaSerValGlyglYlAlaAlaAlaLeuLeuglyLeuAlaAla 192
          ||| ||| |||
Db      572 -----GCTGCTGCGCTTCAAGGCTTCGGTCC 598

RESULT 37
LOCUS   BU065692 611 bp mRNA linear EST 26-AUG-2002
DEFINITION Fgr_8_E01_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA
sequence.
ACCESSION BU065692
VERSION   BU065692.1 GI:22505981
KEYWORDS  EST.
SOURCE    Gibberella zeae
ORGANISM  Gibberella zeae
          Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
          Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE 1 (bases 1 to 611)
AUTHORS   Trail,F., Xu,J.-R., San Miguel,P., Halgren,R.G. and Kistler,H.C.
TITLE     Analysis of expressed sequence tags from Gibberella zeae (anamorph
JOURNAL   Fungal Genet. Biol. 38 (2), 187-197 (2003)
MEDLINE   12508120
PUBMED    12620255
COMMENT   Contact: Frances Trail
          Department of Plant Biology
          Michigan State University
          East Lansing, MI 48824, USA
          Tel: 517 432 2939
          Fax: 517 353 1926
          Email: trail@msu.edu
          Plate: 8 row: E column: 01.
          Location/Qualifiers
FEATURES
     source
         1..611
            /organism="Gibberella zeae"
            /mol_type="mRNA"
            /strain="NRRL 31084"
            /db_xref="taxon:5518"
            /clone_lib="Nitrogen-starved mycelia"
            /note="Vector: Uni-ZapII; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 128 a 229 c 143 g 111 t
ORIGIN
Alignment Scores:
Pred. NO.: 0.00188 Length: 611
Score: 164.50 Matches: 52
Percent Similarity: 40.61% Conservative: 28
Best local Similarity: 26.40% Mismatches: 88
Query Match: 16.08% Indels: 29
DB: 13 Gaps: 7
US-10-081-935-4 (1-194) x BU065692 (1-611)
Qy      1 MetClnPheSerHisAlaLeuIleAlaLeuValAlaAlaGlyLeuAlaSerAlaGlnLeu 20
          ||| ||| |||
Db      80 ATGAGAGTACTCGGCTTGTGCTCTCCCGCTGTGTCCGCTCAGGCTCAGCTCTTT 139
Qy      21 ProAspIleProProCysAlaLeuAsnGlySerpheValGluAlaLeuGlyAsnAspGlyCys 40
          ||| ||| |||
Db      140 GCCACGTCGCCAAGTGGCCATTCCTCTCGCTCGACAGGCCATTCGACGACGACGAC 199

```

```

Qy      41 ThrArgIleuThraSerpheLysCysHisCysSerLys---ProGluLeuProGlyGlnIle 59
          ||| ||| |||
Db      200 TGTGACAGACCGGACCTGCGCTGTGTGTGGAAGGCTTCAACCGCTGTCCATCCAAAGCT 259
          ||| ||| |||
Qy      60 ThrProCysValaGluGluAlaCysProLeuAspAlaArgIleSerValSerAsnIleVal 79
          ||| ||| |||
Db      260 ACCCTTGTGTGTCATTGACGAGTGTGTACCGAC-----GTGCGATTCACGAGAGTGC--- 310
Qy      80 ValAspGlnCysSerLysAlaGlyValProIleAspIleProProValAspThrThrala 99
          ||| ||| |||
Db      311 -----CTTCCCGCTTACCGAACCTCTGC 334
Qy      100 AlaProGluProSerGluThr---AlaGluProThralaGluProthnGluGluProthr 118
          ||| ||| |||
Db      335 AAGAACCCCAAGAGTCTGAGGCCAAGTCCACCCGAG--GAGGAGAACCCACC 391
Qy      119 AlagluProthralagluProthr-----AlagluProthrAlagluProthglu 135
          ||| ||| |||
Db      392 ACCAGCCCGCGCTGCACCTTCGCTGCTTGTCAACACACAGCGCGGCTGTGTAG 451
Qy      136 GluProthralaValProthrglyThrglyglYvalProthrglyThrglySerphe 155
          ||| ||| |||
Db      452 ACCACCCCGCTGCACCACTGCTGCTCCATCTCCACACCGCTGCTGAGAG 511
Qy      156 ThrValThrglyArProthralaSerThProAlagluPheProglYAlaglySerAsn 175
          ||| ||| |||
Db      512 CCGCCACACAGACACCCCGCGCTGCCACCCCTTACCAAGGCTCCGAGACAGCCACGCT 571
Qy      176 ValArgAlaSerValGlyglYlAlaAlaAlaLeuLeuglyLeuAlaAla 192
          ||| ||| |||
Db      572 -----GCTGCTGCGCTTCAAGGCTTCGGTCC 598

RESULT 38
LOCUS   AU284313 612 bp mRNA linear EST 17-SEP-2002
DEFINITION AU284313 Dictyostelium discoideum FC (H. Urushihara) Dictyostelium
discoideum cDNA clone FC-AK17 5', mRNA sequence.
ACCESSION AU284313
VERSION   AU284313.1 GI:23034403
KEYWORDS  EST.
SOURCE    Dictyostelium discoideum
ORGANISM  Dictyostelium discoideum
          Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE 1 (bases 1 to 612)
AUTHORS   Muramoto,T., Suzuki,K., Shimizu,H., Kohara,Y., Koriki,E., Obara,S.,
          Tanaka,Y. and Urushihara,H.
TITLE     Generation and analysis of a gamete-specific gene pool in
          Dictyostelium discoideum (2002b)
JOURNAL   Unpublished
COMMENT   Contact: Hideko Urushihara
          Institute of Biological Sciences
          University of Tsukuba
          1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
          Tel: 81-298-53-4664
          Fax: 81-298-53-6614
          Email: hideko@biol.tsukuba.ac.jp.
          Location/Qualifiers
FEATURES
     source
         1..612
            /organism="Dictyostelium discoideum"
            /mol_type="mRNA"
            /strain="KAX3"
            /db_xref="taxon:44689"
            /clone_lib="FC-AK17"
            /sex="mat A"
            /tissue_type="gamete"
            /dev_stage="Sexual"
            /clone_lib="Dictyostelium discoideum FC (H. Urushihara)"
BASE COUNT 229 a 164 c 62 g 157 t
ORIGIN
Alignment Scores:
Pred. NO.: 0.00206 Length: 612

```


Score:	164.00	Matches:	40
Percent Similarity:	45.26%	Conservative:	3
Best Local Similarity:	42.11%	Mismatch:	48
Query Match:	16.03%	Indels:	4
DB:	9	Gaps:	2

US-10-081-935-4 (1-194) X AU284313 (1-612)

Qy	89	ProIleAspIleProProValAspThrThrAlaIaIaProGluProSerGluThrAlaGlu	108
		:::: ::::	
Db	50	CCAACTGAAACTCCAACTCAAAACACCAACTCAAAACCAACTGAAACTCAAACTCAAACT	109
Qy	109	ProThrAlaGluProThrGluGluProThrAlaGluProThrAlaGluProThrAlaGlu	128
Db	110	CCAACTGAAACTCCAACTGAAACTCCAACTGAAACTCCAACTGAAACTCCAACTGAAACT	169
Qy	129	ProThrHisGluProThrGluGluProThrAlaValProThrGlyThrGlyGlyGlyAla	148
Db	170	CCAACTGAAACTCCAACTGAAACTCCAACTGAAACTCCAACTGAAACTCCAACTGAAACT	229
Qy	149	ProThrGlyThrGlySerPheThrValThrGlyArgProThr-----AlaSerThr	165
		::::	
Db	230	CCAACTGAAACTCCAACTCAAACTCCAACCTGAAACTCCAACCCAAACCCCAACCAACA	289
Qy	166	ProAlaGluPheProGlyAlaGlySerAsnValaGlaGlaSerVal	180
Db	290	CCAAACACCAACCAACCCAGGT---ATTGTGTAATGTTAGTGTT	331

RESULT_39	LOCUS	BU061554	643 bp	mRNA	linear	EST 26-AUG-2002
DEFINITION		Fig_10_D22_T3	Nitrogen-starved	<i>Myxocella</i>	<i>Gibberella</i>	zeae cDNA, mRNA

ACCESSION	BU061554	
VERSION	BU061554.1	GI:225018433
KEYWORDS	EST.	
SOURCE	Gibberella zeae	
ORGANISM	Gibberella zeae	

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
1 (Hance et al., 2002)

REFERENCE	1 (Bases 1 to 643)
AUTHORS	Trall, F., Xu, J.-R., San Miguel, P., Halgren, R.G. and Kistler, H.C.
TITLE	Analysis of expressed sequence tags from <i>Gibberella zeae</i> (anamorphic <i>Fusarium graminearum</i>)
JOURNAL	Fungal Genet. Biol. 38 (2), 187-197 (2003)
LINE	22508120
MEDLINE	12620255
COMMENT	Contact: Frances Trall

COMMENT
Contact: Frances Trail
Department of Plant Biology
Michigan State University
East Lansing, MI 48824, USA
Tel: 517 432 2939
Fax: 517 353 1926
Email: trail@msu.edu
Plate: 10 row: D column: 22

```

FEATURES
    source      Location/Qualifiers
                1..643
                /organism="Gibberella zeae"
                /mol_type="mRNA"
                /strain="NRRL 31084"
                /db_xref="taxon:5518"
                /clone_lib="Nitrogen-starved mycelia"
                /note="Vector: Uni-ZapRI; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT
131 a      240 c      147 g      125 t

```

Alignment Scores:	
Pred. No.:	0.00218
Score:	164.00
Percent Similarity:	42.42%
Best Local Similarity:	26.77%
Query Match:	16.03%
DB:	13
Length:	643
Matches:	53
Conservative:	31
Mismatches:	90
Indels:	24
Gaps:	8

US-10-081-935-4 (1-194) X BU061554 (1-643)

1 MetGlnPheSerHisbaleuilealeuValAlaAlaGlyLeuAlaSerAlaGlnleu 20

.....

Db 150 GCCCAGCTCCCCAAGTGGCCATTCTTGCCTCGACAAGGCCATTGCGCAGGACGACCAGC 207

41 ThrArgLeuThrAspPheLysCysHisCysSerLys---ProGluLeuProGlyGlnIle 59

60 Th-ProCysValGluGluIaIaCysProLeuaspIaArgIleSerValSerAsnIleVal 79

Db 270 ACCTCTTGTCATTTGACGAGTGTGTACCGAC-----GTTGCCATCAACGAGCTC--- 32

D_b 321 -----CTTCCGGCTACCGAGAACCTCTGC 34

oy 100 AlaProGluProSerGluThr---AlaGluProThaAlaGluProThnGluGluProThr 113

119 AlaGluProThrAlaGluProThr-----AlaGluProThrHisGluProThrcgu 133

402AACCAACCGGCGCGACACCTTCACCGACAGCGCGAGCTTGAG 46

Db 462 ACCACCGCGCTGCCACCAACCACTGTGCTTCCATCATTTCCACCAACCGCTGCTGAGAG 521

[illegible]

QY 176 ValArgAlaSerValGlycylIleAlaAlaLeuLeuGlyLeuAlaAlaIleTyr 193

RESULT 40

LOCUS	874 bp	mRNA	linear	EST 03-JUN-2003
CD459301				
DEFINITION	Fg08_03a11 A Fg08 AAFC ECORC Fungarium graminearum complex sub			

ACCESSION	CD459301	GI:31374041
VERSION	CD459301.1	

SOURCE	ORGANISM
Gibberella zeae	Gibberella zeae

REFERENCE
1 (bases 1 to 874)
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

J., Lacroix, C., Quellet, T., Robert, L.S., Singh, J.A., Sproett, D. and Tinker, N.A.

complex plant substrate
unpublished

Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Saskatoon, Saskatchewan S7N 3W9, Canada

CANADA
Tel: (613) 759-1655
Fax: (613) 750-1701

FEATURES
Source 1 874
Email: watsonrj@gr.gc.ca
Location/Qualifiers

/organism="Gibberella zeae"

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 06:44:55 ; Search time 1940 Seconds

(without alignments)
651.460 Million cell updates/sec

Title: US-10-081-935-1

Perfect score: 52
Sequence: 1 atgcagtcctcctcagctct.....cgtcgcgcgcctcgcagctg 52

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	100.0	395	10	BF251279
2	52	100.0	401	10	BF252856
3	52	100.0	531	10	BF251334
4	52	100.0	557	10	BF252723

	5	52	100.0	737	10	BF251103	BF251103	EST148360
	6	39.2	75.4	311	13	B0499065	B0499065	EST08290
	7	32.8	63.1	439	13	B0761304	B0761304	EBR008_SO
	8	32.8	63.1	538	13	B0767128	B0767128	EBR008_SO
	9	32.8	63.1	602	13	B0767421	B0767421	EBR008_SO
	10	29.4	56.5	994	29	CNS06QTV	AL411113	T3 end of
	11	27.4	52.7	559	29	B2235140	B2235140	CH230-477
	12	27.2	52.3	349	14	L38010	L38010	BNAF0232E M
	13	26.8	51.5	1615	12	B1519263	B1519263	603062017
	14	26.6	51.2	559	14	CA766669	CA766669	AF53-RP1
	15	26.6	51.8	850	29	B2564779	B2564779	pac82-164
	16	26.4	50.2	1044	28	A0743420	A0743420	HS 5387 B
	17	26.2	50.4	447	13	B0662115	B0662115	HS071822w
	18	26.2	50.4	456	14	CB883923	CB883923	HS07A16u
	19	26.2	50.4	522	9	AL925722	AL925722	AL925722
	20	26.2	50.4	627	13	B0466638	B0466638	HS01B22T
	21	26.2	50.4	1145	29	AG035039	AG035039	Pan t10g1
	22	26	50.0	885	14	CB680101	CB680101	OSJNEF04G
	23	26	50.0	1057	13	BU504141	BU504141	AGENCOURT
	24	25.8	49.6	353	28	BH225332	BH225332	1006107G1
	25	25.8	49.6	803	29	B2547247	B2547247	OGAKP22TM
	26	25.8	49.6	873	29	B2547241	B2547241	OGAKP22TC
	27	25.6	49.2	123	12	B1189998	B1189998	93d05f8.r
	28	25.6	49.2	478	13	BU667306	BU667306	U002E04 O
	29	25.6	49.2	548	28	BH411461	BH411461	1007022G0
	30	25.6	49.2	630	10	AW984083	AW984083	RC0-HN000
	31	25.6	49.2	773	14	CB657824	CB657824	OSJNEC13H
	32	25.4	48.8	601	14	CA354617	CA354617	626361 NC
	33	25.2	48.5	402	14	CA642822	CA642822	wre1n.pko
	34	25.2	48.5	522	14	CA712672	CA712672	wdk3C.pko
	35	25	48.1	281	14	C72378	C72378	C72378
	36	25	48.1	359	14	CA724779	CA724779	wdf3f.pko
	37	25	48.1	449	9	AW730728	AW730728	CA_Ba02
	38	25	48.1	569	10	BE583376	BE583376	11-9B-MY
	39	25	48.1	591	12	BM370328	BM370328	EBR008_SO
	40	25	48.1	688	14	CB649291	CB649291	OSJNEB13C
	41	25	48.1	755	14	CB648531	CB648531	OSJNEB11P
	42	25	48.1	798	14	CB680316	CB680316	OSJNEF04N
	43	25	48.1	804	10	BF684156	BF684156	963049F02
	44	25	48.1	899	29	B2574472	B2574472	mbh2-3683
	45	25	48.1	1188	29	AG047770	AG047770	Pan t10g1

ALIGNMENTS

RESULT 1
BF251279
LOCUS
DEFINITION
EST148360 Coccidioides immitis spherule cDNA library
immitis cDNA clone CIAAFl 5' sequence, mRNA sequence.
BF251279
ACCESSION
BF251279.1 GI:16931422
VERSION
KEYWORDS
SOURCE
ORGANISM
Coccidioides immitis
Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.
1 (bases 1 to 395)
AUTHORS
TITLE
Gardner, M.J. and Kirkland, T.
JOURNAL
Generation of ESTs from Coccidioides immitis spherule cDNA library
COMMENT
Unpublished
Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@igr.org

FEATURES
source
Location/Qualifiers
1..395
/organism="Coccidioides immitis"
/mol_type="mRNA"

/db_xref="taxon:5501"
/clone="CIAF11"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"
XhoI"
BASE COUNT 78 a 133 c 78 g 106 t
ORIGIN
Query Match 100.0%; Score 52; DB 10; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGCAGTTCTCTCAGCCTCTCATCCCTCTGCTGCGTCCGCGCTCCGACAGTG 52
|||||
142 ATGCAGTTCTCTCAGCCTCTCATCCCTCTGCTGCGTCCGCGCTCCGACAGTG 193
|||||
RESULT 2
BF252856 401 bp mRNA linear EST 15-NOV-2001
LOCUS
DEFINITION EST1420119 Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAB70 5' sequence, mRNA sequence.
ACCESSION BF252856
VERSION BF252856.1 GI:16932999
KEYWORDS
SOURCE
ORGANISM
Coccidioides immitis
Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.
1 (bases 1 to 401)
Gardner, M.J. and Kirkland, T.
Generation of ESTs from Coccidioides immitis spherule cDNA library
Unpublished
Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org.
LOCATION/Qualifiers
FEATURES
source
1..401
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAB70"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 77 a 149 c 89 g 86 t
ORIGIN
Query Match 100.0%; Score 52; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATGCAGTTCTCTCAGCCTCTCATCCCTCTGCTGCGTCCGCGCTCCGACAGTG 52
|||||
59 ATGCAGTTCTCTCAGCCTCTCATCCCTCTGCTGCGTCCGCGCTCCGACAGTG 110
|||||
RESULT 3
BF251334 531 bp mRNA linear EST 15-NOV-2001
LOCUS
DEFINITION EST18594 Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAF68 5' sequence, mRNA sequence.
ACCESSION BF251334
VERSION BF251334.1 GI:16931477
KEYWORDS
EST.

SOURCE
ORGANISM
Coccidioides immitis
Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.
1 (bases 1 to 531)
Gardner, M.J. and Kirkland, T.
Generation of ESTs from Coccidioides immitis spherule cDNA library
Unpublished
Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org.
LOCATION/Qualifiers
FEATURES
source
1..531
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAF68"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 110 a 197 c 117 g 133 t
OY 1 ATGCAGTTCTCTCAGCCTCTCATCCCTCTGCTGCGTCCGCGCTCCGACAGTG 52
|||||
137 ATGCAGTTCTCTCAGCCTCTCATCCCTCTGCTGCGTCCGCGCTCCGACAGTG 188
|||||
RESULT 4
BF252723 557 bp mRNA linear EST 15-NOV-2001
LOCUS
DEFINITION EST14986 Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAA21 5' sequence, mRNA sequence.
ACCESSION BF252723
VERSION BF252723.1 GI:16932866
KEYWORDS
SOURCE
ORGANISM
Coccidioides immitis
Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.
1 (bases 1 to 557)
Gardner, M.J. and Kirkland, T.
Generation of ESTs from Coccidioides immitis spherule cDNA library
Unpublished
Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org.
LOCATION/Qualifiers
FEATURES
source
1..557
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAA21"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_lib="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

ORIGIN

JOURNAL

Query Match 100.0%; Score 52; DB 10; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGTTCTCTCAGCGCTCTCATCGCTCTCGCTGCGCGGCTCGCCAGTG 52
|||||
161 ATGCAGTTCTCTCAGCGCTCTCATCGCTCTCGCTGCGCGGCTCGCCAGTG 212
|||||

Db

RESULT 5
BF251103
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BF251103 737 bp mRNA linear EST 15-NOV-2001
EST1418360 Coccidioides immitis spherule cDNA library Coccidioides
immitis cDNA clone CIAAC67 5' sequence, mRNA sequence.
BF251103
BF251103.1 GI:16931246
EST.
Coccidioides immitis
Coccidioides immitis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.
1 (bases 1 to 737)
Gardner, M.J. and Kirkland, T.
Generation of ESTs from Coccidioides immitis spherule cDNA library
Unpublished
Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@igr.org.

FEATURES
source
1..737
/organism="Coccidioides immitis"
/mol_type="mRNA"
/db_xref="taxon:5501"
/clone="CIAAC67"
/dev_stage="spherule"
/lab_host="SOLR"
/clone_1lb="Coccidioides immitis spherule cDNA library"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 136 a 271 c 168 g 162 t

ORIGIN

Query Match 100.0%; Score 52; DB 10; Length 737;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGTTCTCTCAGCGCTCTCATCGCTCTCGCTGCGCGGCTCGCCAGTG 52
|||||
99 ATGCAGTTCTCTCAGCGCTCTCATCGCTCTCGCTGCGCGGCTCGCCAGTG 150
|||||

Db

RESULT 6
BQ499065
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

BQ499065 311 bp mRNA linear EST 31-OCT-2002
EST08290 PB0001 Paracoccidioides brasiliensis cDNA, mRNA sequence.
BQ499065
BQ499065.1 GI:24452039
EST.
Paracoccidioides brasiliensis
Paracoccidioides brasiliensis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Paracoccidioides.
1 (bases 1 to 311)
Goldman, G.H., Marques, E.R., Ribeiro, D.C.D., Bernardes, L.A., Puccia
R., Travaassos, L.R., Nobrega, F.G., Nobrega, M.P., Savoldi-Barbosa, M.,
Seminhini, C.P. and Goldman, M.H.
The Paracoccidioides brasiliensis EST genome project

BASE COUNT 71 a 169 c 97 g 82 t

ORIGIN

JOURNAL

Query Match 75.4%; Score 39.2; DB 13; Length 311;
Best Local Similarity 84.6%; Pred. No. 0.076;
Matches 44; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGCAGTTCTCTCAGCGCTCTCATCGCTCTCGCTGCGCGGCTCGCCAGTG 52
|||||
91 ATGCAGTTCTCTCAGCGCTCTCATCGCTCTTGTTGCGGCGCAGCAATG 142
|||||

Db

RESULT 7
BQ761304
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ761304 419 bp mRNA linear EST 26-JUL-2002
EBR06_S0001_A05_R root, 3 week, drought-stressed, cv Optic, EBR06
Hordeum vulgare subsp. vulgare cDNA clone EBR06_S0001_A05 5', mRNA
sequence.
BQ761304
BQ761304.1 GI:21969776
EST.
Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 419)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
Development of Barley Transcriptome Resources
Unpublished
Contact: Maugh R. Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: eat@scri.sari.ac.uk.

FEATURES
source
1..419
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/culivar="Optic"
/db_xref="taxon:112509"
/clone="EBR06_S0001_A05"
/cissue_type="root"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_1lb="root, 3 week, drought-stressed, cv Optic,
EBR06"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from roots of 3 week old drought stressed barley
plants. After two weeks in control conditions the seedlings
were raised above the hydroponic solution for 3 hours
daily. N.B. This library has been superseded by EBR06.
Developed as part of the barley transcriptome resources of
BSRC/SEERAD funded cereal IGF (Investigating Gene
Function) project."

BASE COUNT 71 a 169 c 97 g 82 t

ORIGIN

Query Match	Similarity	Score	DB	Length	DB
Best Local	76.9%	Pred. No. 5.8;			
Matches	40;	Conservative	0;	Mismatches	12; Indels 0; Gaps 0;

Db	Query Match	Similarity	Score	DB	Length	DB
27	ATGAAGTTCTCCACCACTCTCATGTCCTTCGTCGCTGCGCGGCTCGCCAGTG	52				

RESULT 8	Db	Query Match	Similarity	Score	DB	Length	DB
B0767128	538 bp	mrna	linear	EST 26-JUL-2002			
LOCUS	EBrc08_S0008_E05_R root. 3 week, drought-stressed, cv Optic, EBrc08						
DEFINITION	Hordeum vulgare subsp. vulgare cDNA clone EBrc08_S0008_E05_5', mRNA						
ACCESSION	B0767128						
VERSION	B0767128.1						
KEYWORDS	EST.						
SOURCE	Hordeum vulgare subsp. vulgare						
ORGANISM	Hordeum vulgare subsp. vulgare						
REFERENCE	Healey, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L., Ramsay, L., Machray, G., Marshall, D.F.M. and Maugh, R.						
AUTHORS	Development of barley 'transcriptome Resources						
TITLE	Unpublished						
JOURNAL	Contact: Waugh R, Marshall DF						
COMMENT	Genome Dynamics/Computational Biology						
	Scottish Crop Research Institute						
	Invergowrie, Dundee, DD2 5DA, Scotland, UK						
	Tel: 00 44 1382 562731						
	Fax: 00 44 1382 562426						
	Email: est@scri.sari.ac.uk.						
FEATURES	Location/Qualifiers						
source	1..538						
	/organism="Hordeum vulgare subsp. vulgare"						
	/mol_type="mRNA"						
	/cultiivar="Optic"						
	/db_xref="taxon:112509"						
	/clone="EBrc08_S0008_E05"						
	/tissue_type="root"						
	/dev_stage="3 week"						
	/lab_host="DH10B"						
	/clone_lib="root, 3 week, drought-stressed, cv Optic, EBrc08"						
	/note="vector: pSPORT1; Site 1: Sal I; Site 2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from roots of 3 week old drought stressed barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERD funded cereal IGF (Investigating Gene Function) project."						

BASE COUNT	ORIGIN
91 a	217 c
119 g	111 t

Query Match	Best Local	Similarity	Score	DB	Length	DB
Matches	40;	Conservative	0;	Mismatches	12; Indels 0; Gaps 0;	

Db	Query Match	Similarity	Score	DB	Length	DB
56	ATGAAGTTCTCCACCACTCTCATGTCCTTCGTCGCTGCGCGGCTCGCCAGTG	52				

RESULT 9	Db	Query Match	Similarity	Score	DB	Length	DB
B0767421	602 bp	mrna	linear	EST 26-JUL-2002			
LOCUS	EBrc08_S0008_122_R root. 3 week, drought-stressed, cv Optic, EBrc08						
DEFINITION	Hordeum vulgare subsp. vulgare cDNA clone EBrc08_S0008_122_5', mRNA						
ACCESSION	B0767421						

VERSION	BQ767421.1		GI:21975895
KEYWORDS	EST.		
SOURCE	Hordeum vulgare subsp. vulgare		
ORGANISM	Hordeum vulgare subsp. vulgare		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae		
	; Triticeae; Hordeum.		
REFERENCE	1 (bases 1 to 602)		
AUTHORS	Hadley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,		
TITLE	Ramsay, L., Machray, G., Marshall, D.F.M. and Maugh, R.		
JOURNAL	Development of Barley Transcriptome Resources		
COMMENT	Unpublished		
	Contact: Maugh R, Marshall DF		
	Genome Dynamics/Computational Biology		
	Scottish Crop Research Institute		
	Invergowrie, Dundee, DD2 5DA, Scotland, UK		
	Tel: 00 44 1382 562731		
	Fax: 00 44 1382 562426		
	Email: est@scri.sari.ac.uk.		
FEATURES	Location/Qualifiers		
source	1..602		
	/organism="Hordeum vulgare subsp. vulgare"		
	/mol_type="mRNA"		
	/catalytic="Optic"		
	/db_xref="taxon:112509"		
	/clone="EBR008_S0008_122"		
	/issue_type="root"		
	/dev_stage="3 week"		
	/lab_host="DH10B"		
	/clone_id="root, 3 week, drought-stressed, cv Optic,		
	EBR008"		
	Note="vector: pSPORT1; Site_1: Sal I; Site_2: Not I;		
	Non-normalised library, directionally cloned into pSPORT1.		
	Derived from roots of 3 week old drought stressed barley		
	plants. Developed as part of the barley transcriptome		
	resources of BBSRC/SEERAD funded cereal ICG (Investigating		
	Gene Function) project."		
BASE COUNT	100 a 251 c 129 g 122 t		
ORIGIN			
Query Match	63.1%; Score 32.8; DB 13; Length 602;		
Best Local Similarity	76.9%; Pred. No. 6.1;		
Matches	40; Conservative 0; Mismatches 12; Indels 0; Gaps 0;		
OR	1 ATGAGTTCTTCACGCTCTCATGCTCTCTGTCGTCGCCGCCCTCGCCAGTG 52		
Db	45 ATGAGTTCTTCACGCTCTGATGCTCTGTTGTCGTCGCCGCCCTCGCCAGCG 96		
RESULT 10			
CNS06QTV			
LOCUS	CNS06QTV 994 bp DNA linear GSS 05-JUN-2001		
DEFINITION	T3 end of clone AW0AA009A12 of library AW0AA from strain CUB 89 of		
ACCESSION	Yarrowia lipolytica, genomic survey sequence.		
VERSION	AL411113		
KEYWORDS	AL411113.1 GI:12180209		
SOURCE	GSS.		
ORGANISM	Yarrowia lipolytica		
	Yarrowia lipolytica		
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
	Saccharomycetales; Dipodascaceae; Yarrowia.		
	1 (bases 1 to 994)		
REFERENCE	Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,		
AUTHORS	Boloitin, F., Bon, E., Broccher, P., Casaregola, S.,		
	DeMontigny, J., Dujon, B., Durieux, P., Lepingle, A., Florente, B.,		
	Malpertuy, A., Neuvéglise, C., Ozier, K., Papadimitrakopoulos, O.,		
	Pottier, S., Saurin, W., Tekala, F., Toffano-Nicolas, C., Wesolowski, L.,		
	Wincker, P. and Weissenbach, J.		
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of		
JOURNAL	yeast species for molecular evolution studies		
MEDLINE	FEMS Lett. 487 (1), 3-12 (2000)		
PUBMED	11152876		

```

REFERENCE
AUTHORS      2 (bases 1 to 994)
              Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C.,
              Artiguenave,F., Wincker,P. and Gaillardin,C.
TITLE        Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
              lipolytica
JOURNAL      FEBS Lett. 487 (1), 95-100 (2000)
MEDLINE      20584727
PUBMED       11152892
REFERENCE    3 (bases 1 to 994)
AUTHORS      Direct Submission
TITLE        Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
              2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
              seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
              This GSS is part of a random genomic sequencing program of thirteen
              yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
              exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
              Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
              lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
              angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
              Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
              5 kb were prepared and both extremities were sequenced. See
              keywords for description of this sequence and for the sequence of
              the other extremity of this insert.
              Location/Qualifiers
FEATURES     source
              1..994
                /organism="Yarrowia lipolytica"
                /mol_type="genomic DNA"
                /strain="CLIB 89"
                /db_xref="taxon:4952"
                /clone="AM0AA009A12"
                /clone_lib="AM0AA"
                /note="end : T3"
                <653..>969
                /note="similar to Saccharomyces cerevisiae ORF YPL123c [
                similarity to ribonucleases ]
                1 putative frameshift(s)
                /evidence=not_experimental
              misc_feature
                234 a 325 c 210 g 221 t 4 others
BASE COUNT   234 a 325 c 210 g 221 t
ORIGIN
Query Match      56.5%; Score 29.4; DB 29; Length 994;
Best Local Similarity 76.6%; Pred. No. 65;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ATGCAGTTCTCTCAGCTCTCATCGCTCTCGTCGCGCGCCCTCGC 47
    |||||
Db 471 ATGCAGTTCTCTCTTGCACCATCGCTACCGTGCCTGCGCGCGCTC 517

RESULT 11
LOCUS      BZ235140
DEFINITION BZ230-477F24.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
ACCESSION  BZ235140
VERSION    BZ235140.1 GI:23893919
KEYWORDS   GSS.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Eukaryota; Chordata; Chordata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE  1 (bases 1 to 559)
AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
              A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
              Jong,P. and Fraser,C.M.
TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL   Unpublished
COMMENT   Other GSSs: CH230-477F24.TV
              Contact: Shaying Zhao
              Department of Eukaryotic Genomics
              The Institute for Genomic Research

2 (bases 1 to 994)
              Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C.,
              Artiguenave,F., Wincker,P. and Gaillardin,C.
TITLE        Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
              lipolytica
JOURNAL      FEBS Lett. 487 (1), 95-100 (2000)
MEDLINE      20584727
PUBMED       11152892
REFERENCE    3 (bases 1 to 994)
AUTHORS      Direct Submission
TITLE        Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
              2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
              seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
              This GSS is part of a random genomic sequencing program of thirteen
              yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
              exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
              Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
              lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
              angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
              Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
              5 kb were prepared and both extremities were sequenced. See
              keywords for description of this sequence and for the sequence of
              the other extremity of this insert.
              Location/Qualifiers
FEATURES     source
              1..994
                /organism="Yarrowia lipolytica"
                /mol_type="genomic DNA"
                /strain="CLIB 89"
                /db_xref="taxon:4952"
                /clone="AM0AA009A12"
                /clone_lib="AM0AA"
                /note="end : T3"
                <653..>969
                /note="similar to Saccharomyces cerevisiae ORF YPL123c [
                similarity to ribonucleases ]
                1 putative frameshift(s)
                /evidence=not_experimental
              misc_feature
                234 a 325 c 210 g 221 t 4 others
BASE COUNT   234 a 325 c 210 g 221 t
ORIGIN
Query Match      56.5%; Score 29.4; DB 29; Length 994;
Best Local Similarity 76.6%; Pred. No. 65;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ATGCAGTTCTCTCAGCTCTCATCGCTCTCGTCGCGCGCCCTCGC 47
    |||||
Db 471 ATGCAGTTCTCTCTTGCACCATCGCTACCGTGCCTGCGCGCGCTC 517

RESULT 11
LOCUS      BZ235140
DEFINITION BZ230-477F24.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
ACCESSION  BZ235140
VERSION    BZ235140.1 GI:23893919
KEYWORDS   GSS.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Eukaryota; Chordata; Chordata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
              Rattus.
REFERENCE  1 (bases 1 to 559)
AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn
              A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de
              Jong,P. and Fraser,C.M.
TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL   Unpublished
COMMENT   Other GSSs: CH230-477F24.TV
              Contact: Shaying Zhao
              Department of Eukaryotic Genomics
              The Institute for Genomic Research

```

```

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
page: http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html
Plate: 477 row: F column: 24
Seq primer: SP6
Class: BAC ends.
FEATURES     source
              Location/Qualifiers
              1..559
                /organism="Rattus norvegicus"
                /mol_type="genomic DNA"
                /strain="BN/SSNHsd/MCW"
                /db_xref="taxon:10116"
                /clone="CH230-477F24"
                /sex="Female"
                /cell_type="Brain"
                /clone_lib="CHORI-230 Segment 2"
                /note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;
                CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by
                Pieter de Jong"
BASE COUNT   144 a 143 c 116 g 156 t
ORIGIN
Query Match      52.7%; Score 27.4; DB 29; Length 559;
Best Local Similarity 75.6%; Pred. No. 2.3e+02;
Matches 34; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 4 CAGTCTCTCAGCTCTCATCGCTCTCGTCGCTCGCGCCCTCGCC 48
    |||||
Db 25 CAGCTCTCTCAGTCTCTCAAGTCTCTTACCTGCTTACCTGCTCACC 69

RESULT 12
LOCUS      L38010
DEFINITION BNAF0225E Mustard flower buds Brassica rapa cdna clone F0225, mRNA
ACCESSION  L38010
VERSION    L38010.1 GI:887186
KEYWORDS   EST.
SOURCE     Brassica rapa
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
              ; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 349)
AUTHORS   Lim,C.O., Kim,H.Y., Kim,M.G., Lee,S.I., Chung,W.S., Park,S.H.,
              Hwang,I. and Cho,M.J.
TITLE     Expressed sequence tags of Chinese cabbage flower bud cdna
JOURNAL   Plant Physiol. 111 (2), 577-588 (1996)
MEDLINE   96255495
PUBMED    8787028
COMMENT   Contact: Lim,C.O., Kim,M.G., Hwang,I. and Cho,M.J.
              Plant Molecular Biology and Biotechnology Research Center
              Email: pmbrc@nongae.gsnu.ac.kr.
              Location/Qualifiers
              1..349
                /organism="Brassica rapa"
                /mol_type="mRNA"
                /strain="pekinensis"
                /db_xref="taxon:3711"
                /clone="F0225"
                /clone_lib="Mustard flower buds"
                /note="Devel_stage = flower bud"
BASE COUNT   49 a 110 c 67 g 123 t
ORIGIN

```



```

/dev_stage="0-16 hours after imbibition"
/lab_host="XL10-Gold"
/clone_lib="HS"
/notes="Vector: pBluescript SK+, Site_1: EcoRI (5'-end of
cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
artifact caused by the kit, in most cases the EcoRI site
is NOT present, as well as the EcoRI adapter used for
cloning. To excise the insert, restriction sites upstream
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
due to the cloning system used Blue/white selection for
recombinants is not 100% reliable."
BASE COUNT      72 a   246 c   216 g   93 t
ORIGIN

```

```

Query Match      50.4%; Score 26.2; DB 13; Length 627;
Best Local Similarity 72.3%; Pred. No. 5.2e+02;
Matches 34; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

Qy 2 TGCAGTCTCTCAGCTCTCATCGCTCTCGTGGTCCGCCGCTCTCC 48
Db 322 TGCAGTCTCTCAGCTCTCATCGCTCTCGTGGTCCGCCGCTCTCC 368

```

```

RESULT 21
AG035039/c
LOCUS
DEFINITION
AG035039
ACCESSION
AG035039.1 GI:16561912
KEYWORDS
SOURCE
ORGANISM

```

```

Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 1145)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpanzee@gsr.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. 1145
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-010J02.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      346 a   173 c   530 g   24 t   72 others
ORIGIN

```

```

Query Match      50.4%; Score 26.2; DB 29; Length 1145;
Best Local Similarity 75.6%; Pred. No. 5.7e+02;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

```

Qy 346 a 173 c 530 g 24 t 72 others

```

```

RESULT 23
BU504141
LOCUS
DEFINITION
AGENCOURT_8968970 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6491918
5', mRNA sequence.
BU504141
ACCESSION
BU504141.1 GI:22810374
KEYWORDS
SOURCE
ORGANISM

```

```

Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Qy 8 TCTCTCAGCTCTCATCGCTCTCGTGGTCCGCCGCTCTCC 48
Db 232 TCINCCCTCTCTCTCGCGCTCTTCGCTGCCNCCCTCTCC 192

```

```

RESULT 22
CB680101
LOCUS
DEFINITION

```

```

OSJNEF04G16.f OSJNEF Oryza sativa (japonica cultivar-group) cDNA
clone OSJNEF04G16 5', mRNA sequence.
CB680101
ACCESSION
CB680101.1 GI:29683826
KEYWORDS
SOURCE
ORGANISM

```

```

Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Magnoliopsida; Euphorbia; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 885)
Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 04 row: G column: 16
Seq primer: gta aaa cga cgg cca gtc.
Location/Qualifiers
1..885
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEF04G16"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEF"
/notes="vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; Uninfected Control"
BASE COUNT      116 a   370 c   249 g   149 t   1 others
ORIGIN

```

```

Query Match      50.0%; Score 26; DB 14; Length 885;
Best Local Similarity 76.2%; Pred. No. 6.2e+02;
Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

```

Qy 9 CTCTCAGCTCTCATCGCTCTCGTGGTCCGCCGCTCTCC 50
Db 384 CGTCCGCTCTCTCTCGCTCTTCGCCGCGCTCTCC 425

```

```

RESULT 23
BU504141
LOCUS
DEFINITION
AGENCOURT_8968970 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6491918
5', mRNA sequence.
BU504141
ACCESSION
BU504141.1 GI:22810374
KEYWORDS
SOURCE
ORGANISM

```

```

Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Query Match	49.6%	Score 25.8;	DB 29;	Length 803;
Best Local Similarity	81.1%	Pred. No. 7e+02;		
Matches 30;	Conservative	0;	Mismatches 7;	Indels 0;
				Gaps 0;

Query Match	49.6%	Score 25.8;	DB 29;	Length 803;
Best Local Similarity	81.1%	Pred. No. 7e+02;		
Matches 30: Conservative	0:	Mismatches 7:	Indels 0:	Gaps 0:

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
BI189998							
BI189998.1	GI:14663677	EST		Fusarium sporotrichioides			
				Fusarium sporotrichioides			
				Fusarium sporotrichioides			
				Eukaryota; Fungi; Ascomycota; Peizomycotina; Sordariomycetes; Hypocorymycetidae; Hypocreales; Mitosporic Hypocreales; Fusarium			
				1 (bases 1 to 123)			
				Ren.Q., Tag,A., Papiow,A., Lai,H., Kupfer,C., Peterson,A., Bereman,M. and Ros,B.			
				Analysis of a Fusarium sporotrichioides EST database			

PERIODS	SOURCE	BASE	COUNT
1960-1969	1960-1969	1960-1969	1960-1969
1970-1979	1970-1979	1970-1979	1970-1979
1980-1989	1980-1989	1980-1989	1980-1989
1990-1999	1990-1999	1990-1999	1990-1999
2000-2009	2000-2009	2000-2009	2000-2009
2010-2019	2010-2019	2010-2019	2010-2019
2020-2029	2020-2029	2020-2029	2020-2029
2030-2039	2030-2039	2030-2039	2030-2039
2040-2049	2040-2049	2040-2049	2040-2049
2050-2059	2050-2059	2050-2059	2050-2059
2060-2069	2060-2069	2060-2069	2060-2069
2070-2079	2070-2079	2070-2079	2070-2079
2080-2089	2080-2089	2080-2089	2080-2089
2090-2099	2090-2099	2090-2099	2090-2099
2100-2109	2100-2109	2100-2109	2100-2109
2110-2119	2110-2119	2110-2119	2110-2119
2120-2129	2120-2129	2120-2129	2120-2129
2130-2139	2130-2139	2130-2139	2130-2139
2140-2149	2140-2149	2140-2149	2140-2149
2150-2159	2150-2159	2150-2159	2150-2159
2160-2169	2160-2169	2160-2169	2160-2169
2170-2179	2170-2179	2170-2179	2170-2179
2180-2189	2180-2189	2180-2189	2180-2189
2190-2199	2190-2199	2190-2199	2190-2199
2200-2209	2200-2209	2200-2209	2200-2209
2210-2219	2210-2219	2210-2219	2210-2219
2220-2229	2220-2229	2220-2229	2220-2229
2230-2239	2230-2239	2230-2239	2230-2239
2240-2249	2240-2249	2240-2249	2240-2249
2250-2259	2250-2259	2250-2259	2250-2259
2260-2269	2260-2269	2260-2269	2260-2269
2270-2279	2270-2279	2270-2279	2270-2279
2280-2289	2280-2289	2280-2289	2280-2289
2290-2299	2290-2299	2290-2299	2290-2299
2300-2309	2300-2309	2300-2309	2300-2309
2310-2319	2310-2319	2310-2319	2310-2319
2320-2329	2320-2329	2320-2329	2320-2329
2330-2339	2330-2339	2330-2339	2330-2339
2340-2349	2340-2349	2340-2349	2340-2349
2350-2359	2350-2359	2350-2359	2350-2359
2360-2369	2360-2369	2360-2369	2360-2369
2370-2379	2370-2379	2370-2379	2370-2379
2380-2389	2380-2389	2380-2389	2380-2389
2390-2399	2390-2399	2390-2399	2390-2399
2400-2409	2400-2409	2400-2409	2400-2409
2410-2419	2410-2419	2410-2419	2410-2419
2420-2429	2420-2429	2420-2429	2420-2429
2430-2439	2430-2439	2430-2439	2430-2439
2440-2449	2440-2449	2440-2449	2440-2449
2450-2459	2450-2459	2450-2459	2450-2459
2460-2469	2460-2469	2460-2469	2460-2469
2470-2479	2470-2479	2470-2479	2470-2479
2480-2489	2480-2489	2480-2489	2480-2489
2490-2499	2490-2499	2490-2499	2490-2499
2500-2509	2500-2509	2500-2509	2500-2509
2510-2519	2510-2519	2510-2519	2510-2519
2520-2529	2520-2529	2520-2529	2520-2529
2530-2539	2530-2539	2530-2539	2530-2539
2540-2549	2540-2549	2540-2549	2540-2549
2550-2559	2550-2559	2550-2559	2550-2559
2560-2569	256		

```

ORIGIN
Query Match          49.2%; Score 25.6; DB 13; Length 478;
Best Local Similarity 77.5%; Pred. No. 7.4e+02;
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 9 CTCTCAGCTCTCATCGCTCTCGTCGCTGCGCGGCTCGCC 48
Db 226 CGCTCATGCTCTCATGCGCTCTCGTCGCGCGGCGCGCC 265

RESULT 29
BH411461
LOCUS 1007022G03.x2 1007 - RescueMu Grid H Zea mays genomic, genomic
DEFINITION survey sequence.
ACCESSION BH411461
VERSION BH411461.1 GI:17582978
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 548)
AUTHORS Walbot,V.
TITLE Maize genomic sequences found using engineered RescueMu transposon
JOURNAL Unpublished
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 1007022 column: 11
Class: transposon-tagged.
FEATURES
source
location/Qualifiers
1..548
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A18/B73"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="1007 - RescueMu Grid H"
/note="Organ: leaf; Vector: RescueMu (engineered from
pBlueScript backbone); Site 1: BamHI; Site 2: BglII;
RescueMu is a 4.9 kb, modified maize Mu transposon
designed to allow plasmid rescue from total genomic DNA.
Mu elements insert preferentially into transcription
units. For more information on RescueMu, go to the web
site 'www.zmdb.iastate.edu' and follow the links for
'RescueMu.' Grid H was grown at Berkeley in 2001. DNA
was extracted from leaf punches, double digested using
BamHI and BglII, and ligated to form circular plasmids.
DH10B cells were transformed and then screened on LB
plates with ampicillin."
BASE COUNT 100 a 171 c 199 g 78 t
ORIGIN

Query Match          49.2%; Score 25.6; DB 28; Length 548;
Best Local Similarity 70.8%; Pred. No. 7.6e+02;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 2 TGCAGTTCTCTCAGCTCTCATCGCTCTCGTCGCTGCGCGGCTCGCCA 49
Db 364 TGCAGCTCCACCGCCATCGCTCTCGTCGCTGCGCGGCGCGGCCA 411

RESULT 30
AW984083/c

```

```

LOCUS
DEFINITION RC0-HN00006-160300-011-h09 HN0006 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW984083
VERSION AW984083.1 GI:8175681
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 630)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,C.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-RC0-HN0006-160
300-011-h09&t3=2000-03-16&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 5
High quality sequence stop: 478.
FEATURES
source
location/Qualifiers
1..630
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HN0006"
/note="Organ: head normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 148 a 192 c 187 g 103 t
ORIGIN

Query Match          49.2%; Score 25.6; DB 10; Length 630;
Best Local Similarity 70.8%; Pred. No. 7.7e+02;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 4 CAGTTCTCTCAGCTCTCATCGCTCTCGTCGCTGCGCGGCTCGCCAGT 51
||| ||||| ||| ||| ||||| ||||| ||||| |||||
Db 234 CACTTGTCTACCGCGTGCAGCATCCGCTCCCTGCGCGGCTCGCCGT 187
||| ||||| ||| ||| ||||| ||||| ||||| |||||

RESULT 31
CB657824
LOCUS OSJNEC13H04.f OSJNEC Oryza sativa (japonica cultivar-group) cDNA
DEFINITION clone OSJNEC13H04 5', mRNA sequence.
ACCESSION CB657824
VERSION CB657824.1 GI:29661549
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

REFERENCE AUTHORS

1 (bases 1 to 773)
Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E., Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.

TITLE

Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea

JOURNAL COMMENT

Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu

PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g

Plate: 13 row: H column: 04
Seq primer: gta aaa cga cgg cca gtc.

FEATURES source

1..773
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/clones="OSJNEC13H04"

/tissue type="Leaf"

/dev stage="3 week"

/lab_host="DH10B"

/clone_lib="OSJNEC"

/notes="vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"

BASE COUNT 139 a 265 c 240 g 129 t

ORIGIN

Query Match 49.2%; Score 25.6; DB 14; Length 773;

Best Local Similarity 77.5%; Pred. No. 8e+02;

Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 9 CTCTCAGCTCTCATCGCTCTCGTGGCGGCGCTCGCC 48

Db 173 CGCTCATGCTCTCATCGCTCTCGTGGCGGCGGCGGCGCC 212

RESULT 32

CA354617

LOCUS 601 bp mRNA linear EST 05-NOV-2002

DEFINITION 626361 NCCCA IRT Oncorhynchus mykiss cDNA clone IRT78D19_B10 5', mRNA sequence.

ACCESSION CA354617.1 GI:24599804

VERSION CA354617

KEYWORDS EST.

SOURCE Oncorhynchus mykiss (rainbow trout)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei;

Proactinopterygii; Salmoniformes; Salmonidae; Oncorhynchus. 1 (bases 1 to 601)

REFERENCE Rexroad, C.E. and Keele, J.W.

AUTHORS Sequence analysis of a rainbow trout normalized cDNA library

TITLE Unpublished

JOURNAL Contact: Rexroad CE

COMMENT USDA, ARS, National Center for Cool and Cold Water Aquaculture

11876 Leetown Road, Kearneysville, WV 25430, USA

Tel: 304 724 8340 x2129

Fax: 304 725 0351

Email: crexroad@nccca.ars.usda.gov

Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified by cross match v0.990329.

Seq primer: AGCGGATACAAATTCACACGGA.

FEATURES source

1..601
Location/Qualifiers

/organism="Oncorhynchus mykiss"

/mol_type="mRNA"

/db_xref="taxon:9022"

/clones="IRT78D19_B10"

/tissue type="pooled"

/lab_host="DH10B"

/clone_lib="NCCCA IRT"

/notes="vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from brain, gill, liver, spleen, muscle, and kidney."

BASE COUNT 132 a 166 c 131 g 172 t

ORIGIN

Query Match 48.8%; Score 25.4; DB 14; Length 601;

Best Local Similarity 68.6%; Pred. No. 8.8e+02;

Matches 35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 TGCAGTTCTCTCAGCTCTCATCGCTCTCGTGGCGGCTCGCCAGTG 52

Db 266 TGCTTTGCTCTCACACTTACCTCTCTTCCAGGGCGGCATGACTGTG 316

RESULT 33

CA642822

LOCUS 402 bp mRNA linear EST 23-NOV-2002

DEFINITION wreln.pk0061.b12 wreln Triticum aestivum cDNA clone

ACCESSION CA642822

VERSION CA642822.1 GI:25221118

KEYWORDS EST.

SOURCE Triticum aestivum (bread wheat)

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

REFERENCE 1 (bases 1 to 402)

AUTHORS Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z., Miao, G., Caraher, N. and Hanafey, M.K.

TITLE DuPont wheat cDNA Sequence

JOURNAL Unpublished

COMMENT Contact: Scott V. Tingey

Crop Genetics

E. I. DuPont de Nemours and Company

1 Innovation Way, P.O. Box 6104, Newark, DE 13714-6104, USA

Tel: 302-631-2602

Fax: 302-631-2607

Email: Scott.V.Tingey@USA.dupont.com

Seq primer: M13.

FEATURES

source

1..402
Location/Qualifiers

/organism="Triticum aestivum"

/mol_type="mRNA"

/db_xref="taxon:4565"

/clones="wreln.pk0061.b12"

/tissue type="root"

/clone_lib="wreln"

/notes="vector: pBluescript SK+; Site 1: EcoRI; Site 2: XhoI; Wheat (Triticum aestivum L.) root; normalized from wreln library"

BASE COUNT 126 a 85 c 106 g 69 t 16 others

ORIGIN

Query Match 48.5%; Score 25.2; DB 14; Length 402;

Best Local Similarity 75.0%; Pred. No. 9.5e+02;

Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 7 TTCTCTCAGCTCTCATCGCTCTCGTGGCGGCGCTCG 46

Db 210 TTCTTTNCTTCGACCATCGCTCTCTTCTTGTGCGGCTCG 171

```
RESULT 34
CA712672      522 bp  mRNA  linear  EST 26-NOV-2002
LOCUS      wdk3c.pk006.l16 wdk3c Triticum aestivum cDNA clone wdk3c.pk006.l16
DEFINITION      5' end, mRNA sequence.
ACCESSION      CA712672
VERSION      CA712672.1 GI:25434465
KEYWORDS      EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Triticum.
REFERENCE      1 (bases 1 to 522)
AUTHORS      Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
TITLE      DuPont Wheat cDNA Sequence
JOURNAL      Unpublished
COMMENT      Contact: Scott V. Tingey
Crop Genetics
E. I. Dupont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
Location/Qualifiers
1..522
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wdk3c.pk006.l16"
/tissue_type="kernel"
/lab_host="DH10B"
/clone_lib="wdk3c"
/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum L.) developing kernel, 14
days after anthesis."
BASE COUNT      58 a 243 c 138 g 80 t 3 others
ORIGIN
Query Match      48.5%; Score 25.2; DB 14; Length 522;
Best Local Similarity 71.7%; Pred. No. 9.8e-02;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy      3 GCAGTTCTCTCAGCGTCTCATCGCTCTCGCTCGCGCGCTCGCC 48
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      276 GCAGTGGGCTCCCTTCTCTCGCGCGCTCATCGCGCGCGCTCGCC 321

RESULT 35
C72378
LOCUS      C72378 Rice panicle at flowering stage Oryza sativa
DEFINITION      cultivar-group) cDNA clone E1517_2A, mRNA sequence.
ACCESSION      C72378
VERSION      C72378.1 GI:2427915
KEYWORDS      EST.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM      Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE      1 (bases 1 to 281)
AUTHORS      Sasaki,T. and Yamamoto,K.
TITLE      Rice cDNA from panicle at flowering stage
JOURNAL      Unpublished
COMMENT      Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
```

```
Email: teasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
PROJECT='RGP'.
Location/Qualifiers
1..281
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="E1517_2A"
/dev_stage="flowering stage"
/clone_lib="Rice panicle at flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering
stage"
BASE COUNT      45 a 98 c 96 g 39 t 3 others
ORIGIN
Query Match      48.1%; Score 25; DB 14; Length 281;
Best Local Similarity 69.4%; Pred. No. 1.e+03;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy      1 ATGCAGTTCTCTCAGCGTCTCATCGCTCTCGCTCGCGCGCTCGCCA 49
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      62 ATGCAGCGGGCTCTCGCGCTCGCTCGCTCTCGCGCGCGCGCGCCA 110

RESULT 36
CA724779      359 bp  mRNA  linear  EST 26-NOV-2002
LOCUS      wds3f.pk001.f15 wds3f Triticum aestivum cDNA clone wds3f.pk001.f15
DEFINITION      5' end, mRNA sequence.
ACCESSION      CA724779
VERSION      CA724779.1 GI:25446575
KEYWORDS      EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Triticum.
REFERENCE      1 (bases 1 to 359)
AUTHORS      Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
TITLE      DuPont Wheat cDNA Sequence
JOURNAL      Unpublished
COMMENT      Contact: Scott V. Tingey
Crop Genetics
E. I. Dupont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
Location/Qualifiers
1..359
/organism="Triticum aestivum"
/db_xref="taxon:4565"
/clone="wds3f.pk001.f15"
/tissue_type="seedling"
/lab_host="DH10B"
/clone_lib="wds3f"
/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat 10 day old seedling full length without
biotin"
BASE COUNT      75 a 134 c 60 g 75 t 15 others
ORIGIN
Query Match      48.1%; Score 25; DB 14; Length 359;
Best Local Similarity 69.4%; Pred. No. 1.1e+03;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy      1 ATGCAGTTCTCTCAGCGTCTCATCGCTCTCGCTCGCGCGCTCGCCA 49
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      80 ATGGTTTTCTCCACCATTGCCACCGCTGTATCGCGCGCGCGTGTGCCA 128
```


RESULT 37
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AW730728 449 bp mRNA linear EST 08-MAR-2001
GA_Ea027103 Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ea027103, mRNA sequence.

AW730728
AW730728.1 GI:7628277
EST.

Gossypium arboreum
Gossypium arboreum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.

REFERENCE

AUTHORS
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A., and Wilkins, T.A.

TITLE

An integrated analysis of the genetics, development, and evolution
of the cotton fiber

JOURNAL

COMMENT
Unpublished
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

High quality sequence stop: 449.

FEATURES

source

BASE COUNT
ORIGIN

127 a 185 c 170 g 87 t

/organism="Gossypium arboreum"

/mol_type="mRNA"

/strain="AKA"

/cultivar="8400"

/db_xref="taxon:29729"

/clone="GA_Ea027103"

/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"

/lab_host="E. coli"

/clone_lib="Gossypium arboreum 7-10 dpa fiber library"

/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT
ORIGIN

115 a 78 c 184 g 72 t

Query Match

Best Local Similarity 48.1%; Score 25; DB 9; Length 449;

Mismatches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY

2 TCGAGTTCTCAGCGTCTCATCGCTCTCGTCGTCGCGCGGC 42

DB 262 TTTCGGTTTCGACGCTCTCATCGCTCTCGTCGTCGCGGC 222

RESULT 38
BE583376
LOCUS

DEFINITION
11-9B-MY Paojaemy Phytophthora sojae cDNA, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE583376
BE583376.1 GI:9834325
EST.

Phytophthora sojae

Phytophthora sojae

Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.

1 (bases 1 to 569)

Qutob, D., Hraber, P.T., Sobral, B.W.S. and Gijsen, M.

Comparative analysis of expressed sequences in Phytophthora sojae

Plant Physiol. 123 (1), 243-254 (2000)

20267956

10806241

CONTACT: Gijsen M

Agriculture and Agri-Food Canada

FEATURES
source

Seq primer: M13 reverse.
Location/Qualifiers
1..591
/organism="Hordeum vulgare subsp. vulgare"

1..591

1391 Sandford Street, London, Ontario, Canada N5V 4T3
Tel: 519 457 1470
Fax: 519 457 3997
Email: gijzen@em.agr.ca.

FEATURES
source

Location/Qualifiers
1..569

/organism="Phytophthora sojae"

/mol_type="mRNA"

/strain="race 2, strain P6497"

/db_xref="taxon:67593"

/dev_stage="mycelium"

/lab_host="E. coli strain XL0LR"

/clone_lib="Psojaemy"

/note="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI. This
cDNA library was constructed from polyA+ enriched mRNA
from mycelium grown in liquid medium for 3 weeks at 25 C
in the dark in a liquid synthetic medium of 2.4 g sucrose
, 0.27 g asparagine, 10 mg cholesterol, 10 mg ascorbate,
and 2 mg thiamine, per liter, plus salts and minerals.

Complementary DNA was synthesized from mRNA using an
XhoI-poly(dT) linker-primer. EcoRI adapters were ligated
to the blunt-ended cDNA fragments and the products were
digested with XhoI for directional cloning into lambda ZAP
Express vector. This lambda library was amplified once
using E. coli host strain XL1 Blue MRF. Inserts were then
subcloned by mass excision using ExAssist helper phage for
conversion into phagemid vector pBK-CMV in E. coli host
strain XL0LR. Sequenced using T3 primer: 5' ATT AAC CCT
CAC TAA AGG GA 3'."

BASE COUNT 127 a 185 c 170 g 87 t

ORIGIN

Query Match 48.1%; Score 25; DB 10; Length 569;

Best Local Similarity 69.4%; Pred. No. 1.1e+03;

Mismatches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY

1 ATGCGAGTTCTCAGCGTCTCATCGCTCTCGTCGTCGCGCGCTCGCCA 49

DB 44 ATGAAGCTCTGCTGCTGCTCTCGCTCGCGCGCGCTTGGCCA 92

RESULT 39
BM370328
LOCUS

DEFINITION

BM370328 591 bp mRNA linear EST 23-JUL-2002
EBro08_SQ003_N19_R root, 3 week, drought-stressed, cv Optic, EBro08
Hordeum vulgare subsp. vulgare cDNA clone EBro08_SQ003_N19_5', mRNA
sequence.

BM370328

BM370328.2 GI:21951439

EST.

Hordeum vulgare subsp. vulgare

Hordeum vulgare subsp. vulgare

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.

1 (bases 1 to 591)

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
Development of Barley Transcriptome Resources
Unpublished
On Jan 10, 2002 this sequence version replaced gi:18113718.

Contact: Waugh R, Marshall DF

Genome Dynamics/Computational Biology

Scottish Crop Research Institute

Invergowrie, Dundee, DD2 5DA, Scotland, UK

Tel: 00 44 1382 562731

Fax: 00 44 1382 562426

Email: est@scri.sari.ac.uk

All sequence has a Phred quality score of 20 or over

Seq primer: M13 reverse.

Location/Qualifiers

1..591

/organism="Hordeum vulgare subsp. vulgare"

/mol_type="mRNA"
/cultivar="Optic"
/db_xref="taxon:112509"
/clone="EBro08_SQ003_N19"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="root, 3 week, drought-stressed, cv Optic,
EBro08"
/note="vector: pSPORT1; Site 1: Sal I; Site 2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from roots of 3 week old drought stressed barley
plants. Developed as part of the barley transcritome
resources of BBSRC/SEERAD funded cereal IGF (Investigating
Gene Function) project."
BASE COUNT 114 a 248 c 108 g 121 t
ORIGIN

Query Match 48.1%; Score 25; DB 12; Length 591;
Best Local Similarity 69.4%; Pred. No. 1.1e+03;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 ATGCAGTTCTCTCAGCTCTCATCGCTCTCGTCCGCGCTCGGCA 49
||||| | | | | | | | | | | | | | | | | | | | | |
Db 84 ATGCAGTTCCGACGCGCTCTTGCCTCGCGCTCGGCGTTGCGCA 132

RESULT 40
CB649291 688 bp mRNA linear EST 08-APR-2003
LOCUS OSJNEB13C08.f OSJNEB Oryza sativa (japonica cultivar-group) cDNA
DEFINITION clone OSJNEB13C08 5', mRNA sequence.
CB649291
ACCESSION CB649291 GI:29644284
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehartoideae; Oryzaceae; Oryza.
1 (bases 1 to 688)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 13 row: C column: 08
Seq primer: gta aaa cga cgg cca gtc.
Location/Qualifiers
1 .688
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEB13C08"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEB"
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
)"

BASE COUNT 98 a 264 c 221 g 105 t
ORIGIN

Query Match 48.1%; Score 25; DB 14; Length 688;
Best Local Similarity 69.4%; Pred. No. 1.2e+03;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 ATGCAGTTCTCTCAGCTCTCATCGCTCTCGTCCGCGCTCGGCA 49
||||| | | | | | | | | | | | | | | | | | | | | |
Db 70 ATGCAGCGGGCTCTCGCGCTCTCGCGCTCTCGCGCGCGCGCA 118

Search completed: November 13, 2003, 08:51:58
Job time : 1949 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 01:09:54 ; Search time 1945 Seconds
(without alignments)
1093.728 Million cell updates/sec

Title: US-10-081-935-1

Perfect score: 52

Sequence: 1 atgcagcttctcaagctctt.....cgctgcggcctgcacagtg 52

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.ste.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	52	100.0	1201	8	CIU39835	U39835 Coccidioide
2	52	100.0	1234	8	CIU32518	U32518 Coccidioide
3	52	100.0	1435	8	CIU51200	U51200 Coccidioide
4	52	100.0	3821	8	AF013256	AF013256 Coccidioi
5	28.2	54.2	197059	2	AC105307	AC105307 Bos tauru
6	27.4	52.7	219080	2	AC120087	AC120087 Rattus no
7	27.4	52.7	253253	2	AC126527	AC126527 Rattus no
8	26.8	51.5	55939	2	AC091017	AC091017 Homo sapi
9	26.6	51.2	148301	8	AC098566	AC098566 Oryza sat
10	26.6	51.2	306067	8	AE017063	AE017063 Oryza sat
11	26.4	50.8	3582	8	AF092435	AF092435 Phaeospha
12	26.2	50.4	135357	2	AP004769	AP004769 Oryza sat
13	26.2	50.4	230357	2	AC140331	AC140331 Mus muscu
14	26	50.0	2595	3	LEISERTHRE	LEIS559 Leishmania
15	26	50.0	117846	2	AC073431	AC073431 Homo sapi
16	26	50.0	149082	2	AP005635	AP005635 Oryza sat
17	26	50.0	162208	2	AP005637	AP005637 Oryza sat
18	26	50.0	171665	2	AC084081	AC084081 Homo sapi
19	26	50.0	187336	2	AP005427	AP005427 Oryza sat
20	26	50.0	211291	9	AC107959	AC107959 Homo sapi
21	25.6	49.2	6000	4	OCU18102	Y18102 Oryctolagus
22	25.6	49.2	10951	1	AE008035	AE008035 Agrobacte
23	25.6	49.2	11764	1	AE009069	AE009069 Agrobacte
24	25.6	49.2	106732	8	AP005454	AP005454 Oryza sat
25	25.6	49.2	139487	2	AP003513	AP003513 Oryza sat
26	25.6	49.2	164978	2	AP003569	AP003569 Oryza sat
27	25.4	48.8	118444	2	AC114896	AC114896 Oryza sat
28	25.4	48.8	140556	2	AC130601	AC130601 Oryza sat
29	25.4	48.8	144077	2	AC099041	AC099041 Oryza sat
30	25.4	48.8	152359	2	CNS08C87	AL731758 Oryza sat
31	25.4	48.8	168864	2	CNS08CAF	AL831802 Oryza sat
32	25.4	48.8	292200	1	SC0939129	AL939129 Streptomy
33	25.2	48.5	10037	1	AE005738	AE005738 Caulobact
34	25.2	48.5	141983	8	AP003047	AP003047 Oryza sat
35	25.2	48.5	281000	3	TBBCHRIA2	AL929604 Trypanoso
36	25	48.1	723	1	PSU96179	U96179 Pseudomonas
37	25	48.1	2898	1	AF232003	AF232003 Pseudomon
38	25	48.1	3076	1	AF461561	AF461561 Pseudomon
39	25	48.1	3719	1	MKFVCB	X98916 M.kandleri
40	25	48.1	10264	1	AE010324	AE010324 Methanopy
41	25	48.1	11867	1	AE005119	AE005119 Halobacte
42	25	48.1	14623	1	AE004888	AE004888 Pseudomon
43	25	48.1	101237	2	AC026810	AC026810 Homo sapi
44	25	48.1	114354	8	OSUN00227	AL731580 Oryza sat
45	25	48.1	117963	9	AC016651	AC016651 Homo sapi

ALIGNMENTS

RESULT 1
CIU39835
LOCUS
DEFINITION
Coccidioideis immitis immunoreactive cell wall protein mRNA,
complete cds.
U39835
U39835.1 GI:1161373
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Coccidioideis posadasii
Coccidioideis posadasii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioideis.
1 (bases 1 to 1201)
REFERENCE
AUTHORS
Dugger,K.O., Villareal,K.M., Ngyuen,A., Zimmermann,C.R., Law,J.H.
and Galgiani,J.N.

CIU39835 1201 bp mRNA linear PLN 18-JAN-1996
Coccidioideis immitis immunoreactive cell wall protein mRNA,
complete cds.


```
Db 175 ATGCAGTTCTCAGCTCTCATCGCTCTCGTGTGCTGCCGCGCTGCCAGTG 226

RESULT 4
AF013256
LOCUS
DEFINITION Coccidioides immitis proline rich antigen (PRA) gene, complete cds.
ACCESSION AF013256
VERSION AF013256.1 GI:2331288
KEYWORDS
SOURCE
ORGANISM Coccidioides posadasii
Coccidioides posadasii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; Mitosporic Onygenales; Coccidioides.
REFERENCE
1 (bases 1 to 3821)
Peng, T., Orsborn, K.I., Orbach, M.J. and Galgiani, J.N.
Proline-rich vaccine candidate antigen of Coccidioides immitis:
conservation among isolates and differential expression with
spherule maturation
J. Infect. Dis. 179 (2), 518-521 (1999)
99094977
PUBMED 9878042
2 (bases 1 to 3821)
Peng, T., Orsborn, K.I., Orbach, M.J. and Galgiani, J.N.
Direct Submission
Submitted (09-JUL-1997) Infectious Disease, University of Arizona,
1501 N. Campbell, Tucson, AZ 85724, USA
JOURNAL
LOCATION/Qualifiers
FEATURES
1..3821
/organism="Coccidioides posadasii"
/mol_type="genomic DNA"
/db_xref="taxon:199306"
737..2126
/gene="PRA"
Join(737..976,1055..1193,1296..2126)
/gene="PRA"
/product="proline rich antigen"
Join(896..976,1055..1193,1296..1660)
/gene="PRA"
/note="antigen 2"
/codon_start=1
/product="proline rich antigen"
/db_xref="GI:2331289"
/protein_id="AAB66894.1"
/translation="WQFSGHALIALVAAGLASAQLDIPPCALNCFVEALGNDGCTRLT
DFKCHSRPELPQITPCVEEACPLDARISVNIIVDOCSKAGVIDIPPVDTTAAPE
PSETAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPTAEPT
PTASTPAEPFGAGSNRASVGIAALGLAAYL"
BASE COUNT 924 a 981 c 903 g 1013 t
ORIGIN
Query Match 100.0%; Score 52; DB 8; Length 3821;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAGTTCTCAGCTCTCATCGCTCTCGTGTGCTGCCGCGCTGCCAGTG 52
|||||
896 ATGCAGTTCTCAGCTCTCATCGCTCTCGTGTGCTGCCGCGCTGCCAGTG 947
|||||

RESULT 5
AC105307
LOCUS
DEFINITION Bos taurus clone rp42-407c20, WORKING DRAFT SEQUENCE, 8 ordered
pieces.
ACCESSION AC105307
VERSION AC105307.12 GI:27764694
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Bos taurus (cow)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactylia; Ruminantia; Pecora; Bovidae;
```

```

Bovidae; Bovinae; Bos.
1 (bases 1 to 197059)
Zhu, H., Lewin, H.A. and Roe, B.A.
Bos Taurus BAC Clone rp42-407c20
Unpublished
REFERENCE
2 (bases 1 to 197059)
Zhu, H., Lewin, H.A. and Roe, B.A.
Direct Submission
Submitted (29-DEC-2001) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 197059)
Zhu, H., Lewin, H.A. and Roe, B.A.
Direct Submission
Submitted (23-JAN-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Jan 16, 2003 this sequence version replaced gi:23266349.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UONOR
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 3409: contig of 3409 bp in length
* 3410 3509: gap of unknown length
* 3510 9386: contig of 5877 bp in length
* 9387 9486: gap of unknown length
* 9487 19601: contig of 10115 bp in length
* 19602 19701: gap of unknown length
* 19702 37667: contig of 17966 bp in length
* 37668 37767: gap of unknown length
* 37768 55040: contig of 17273 bp in length
* 55041 81739: contig of 26599 bp in length
* 81740 81839: gap of unknown length
* 81840 107841: contig of 25802 bp in length
* 107842 107741: gap of unknown length
* 107742 197059: contig of 89318 bp in length.
FEATURES
Location/Qualifiers
1..197059
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="rp42-407c20"
/clone_lib="RPCI - 42 Male Bovine BAC Library"
BASE COUNT 39457 a 60390 c 59374 g 37137 t 701 others
ORIGIN
Query Match 54.2%; Score 28.2; DB 2; Length 197059;
Best Local Similarity 80.5%; Pred. No. 1.8e+02;
Matches 33; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 8 TCTCTCAGCTCTCATCGCTCTCGTGTGCTGCCGCGCTGCC 48
|||||
Db 182200 TCGTCCCGCTCGCGCGCTGTGCTGCTGCTCTTGCC 182240
|||||

RESULT 6
AC120087/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-345D21, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
ACCESSION AC120087
VERSION AC120087.4 GI:23194658
```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 253253)

REFERENCE AUTHORS

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Fallg, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabiati, A., Gant, R., Garcia, A., Garner, I., Garza, M., Gebregregis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowitz, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulsged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankemelame, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE JOURNAL

2 (bases 1 to 253253)

REFERENCE AUTHORS

Worley, K. C.
Submitted (06-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE AUTHORS

3 (bases 1 to 253253)

TITLE JOURNAL

Rat Genome Sequencing Consortium.

REFERENCE AUTHORS

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:24819062. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GMEX

Center clone name: CH230-157M5

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 223138 bases at least Q40

Consensus quality: 225720 bases at least Q30

Consensus quality: 226900 bases at least Q20

Estimated insert size: 230245; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 250704: contig of 250704 bp in length
* 250705 250804: gap of unknown length
* 250805 251997: contig of 1193 bp in length
* 251998 252097: gap of unknown length
* 252098 253253: contig of 1156 bp in length.

----- Location/Qualifiers

source 1..253253

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-157M5"

1..1369

/note="wgs_end_extension"

clone_end:T7"

3331..4610

/note="wgs_end_extension"

clone_end:T7"

9685..10077

/note="clone_boundary"

clone_end:T7"

site:ECORI

end_sequence:BH286818"

complement(243837..244333)

/note="clone_boundary"

clone_end:Sp6"

site:ECORI

end_sequence:BH286821"

246750..248811

/note="wgs_end_extension"

clone_end:Sp6"

248862..250704

/note="wgs_end_extension"

clone_end:Sp6"

BASE COUNT 64432 a 49523 c 48940 g 64796 t 25562 others

ORIGIN

Query Match 52.7%; Score 27.4; DB 2; Length 253253;

Best Local Similarity 75.6%; Pred. No. 3e+02;

Matches 34; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```

Qy      4  CAGTCTCTCAGCTCTCATCGTCTCGTGGTGGCGGCTCGCC 48
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      227415 CAGCTCTCTCAGCTCTCACTTCTCAAGTCTCTACCTGCTCACC 227459

RESULT 8
AC091017      55939 bp      DNA      linear      HTG 24-MAR-2001
LOCUS      Homo sapiens chromosome 17 clone RP11-64A4 map 17, LOW-PASS
DEFINITION      SEQUENCE SAMPLING.
ACCESSION      AC091017
VERSION      1 GI:13443169
KEYWORDS      HTG; HTGS_PHASE0.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 55939)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamarez,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Riese,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
            Severy,P., Sougniez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (24-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L11979
            Center clone name: 64_A_4
            -----
            * NOTE: This record contains 67 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *
            * 1 740: contig of 740 bp in length
            * 741 840: gap of 100 bp
            * 841 1579: contig of 739 bp in length
            * 1580 1679: gap of 100 bp
            *
            * 1680 2414: contig of 735 bp in length
            * 2415 2514: gap of 100 bp
            * 2515 3262: contig of 748 bp in length
            * 3263 3362: gap of 100 bp
            * 3363 4072: contig of 710 bp in length
            * 4073 4172: gap of 100 bp
            * 4173 4912: contig of 740 bp in length
            * 4913 5012: gap of 100 bp
            * 5013 5769: contig of 757 bp in length
            * 5770 5869: gap of 100 bp
            * 5870 6612: contig of 743 bp in length
            * 6613 6712: gap of 100 bp
            * 6713 7460: contig of 748 bp in length
            * 7461 7560: gap of 100 bp
            * 7561 8321: contig of 760 bp in length
            * 8321 8420: gap of 100 bp
            * 8421 9171: contig of 751 bp in length
            * 9172 9271: gap of 100 bp
            * 9272 9995: contig of 724 bp in length
            * 9996 10095: gap of 100 bp
            * 10096 10862: contig of 767 bp in length
            * 10863 10962: gap of 100 bp
            * 10963 11722: contig of 760 bp in length
            * 11723 11822: gap of 100 bp
            * 11823 12538: contig of 716 bp in length
            * 12539 12638: gap of 100 bp
            * 12639 13378: contig of 740 bp in length
            * 13379 13478: gap of 100 bp
            * 13479 14244: contig of 766 bp in length
            * 14245 14344: gap of 100 bp
            * 14345 15104: contig of 760 bp in length
            * 15105 15204: gap of 100 bp
            * 15205 15920: contig of 716 bp in length
            * 15921 16020: gap of 100 bp
            * 16021 16746: contig of 726 bp in length
            * 16747 16846: gap of 100 bp
            * 16847 17576: contig of 730 bp in length
            * 17577 17676: gap of 100 bp
            * 17677 18394: contig of 718 bp in length
            * 18395 18494: gap of 100 bp
            * 18495 19206: contig of 712 bp in length
            * 19207 19306: gap of 100 bp
            * 19307 20047: contig of 741 bp in length
            * 20048 20474: gap of 100 bp
            * 20475 20910: contig of 763 bp in length
            * 20911 21010: gap of 100 bp
            * 21011 21763: contig of 753 bp in length
            * 21764 21863: gap of 100 bp
            * 21864 22595: contig of 732 bp in length
            * 22596 22696: gap of 100 bp
            * 22697 23425: gap of 100 bp
            * 23426 24250: contig of 726 bp in length
            * 24251 24350: gap of 100 bp
            * 24351 25056: contig of 706 bp in length
            * 25057 25156: gap of 100 bp
            * 25157 25893: contig of 737 bp in length
            * 25894 25993: gap of 100 bp
            * 25994 26759: contig of 766 bp in length
            * 26760 26859: gap of 100 bp
            * 26860 27583: contig of 724 bp in length
            * 27584 27684: gap of 100 bp
            * 27685 28426: contig of 742 bp in length
            * 28427 28525: gap of 100 bp
            * 28526 29247: contig of 721 bp in length
            * 29248 29346: gap of 100 bp
            * 29347 30085: contig of 739 bp in length
            * 30086 30185: gap of 100 bp
            * 30186 30907: contig of 721 bp in length
            * 30908 31006: gap of 100 bp
            * 31007 31738: contig of 732 bp in length
            * 31739 31838: gap of 100 bp
            * 31839 32593: contig of 755 bp in length

```


32594 32693: gap of 100 bp
32694 33443: contig of 750 bp in length
33444 33543: gap of 100 bp
33544 34262: contig of 719 bp in length
34263 34362: gap of 100 bp
34363 35110: contig of 748 bp in length
35111 35210: gap of 100 bp
35211 35971: contig of 761 bp in length
35972 36071: gap of 100 bp
36072 36814: contig of 743 bp in length
36815 36914: gap of 100 bp
36915 37663: contig of 749 bp in length
37664 37763: gap of 100 bp
37764 38600: contig of 737 bp in length
38601 39272: contig of 672 bp in length
39273 39373: gap of 100 bp
39374 40106: contig of 734 bp in length
40107 40206: gap of 100 bp
40207 40887: contig of 681 bp in length
40888 40987: gap of 100 bp
40988 41735: contig of 748 bp in length
41736 41835: gap of 100 bp
41836 42586: contig of 751 bp in length
42587 42686: gap of 100 bp
42687 43434: contig of 748 bp in length
43435 43534: gap of 100 bp
43535 44269: contig of 735 bp in length
44270 44369: gap of 100 bp
44370 45088: contig of 719 bp in length
45089 45188: gap of 100 bp
45189 45933: contig of 745 bp in length
45934 46033: gap of 100 bp
46034 46774: contig of 740 bp in length
46774 46873: gap of 100 bp
46874 47596: contig of 723 bp in length
47597 47696: gap of 100 bp
47697 48439: contig of 743 bp in length
48440 48539: gap of 100 bp
48540 49282: contig of 743 bp in length
49283 49382: gap of 100 bp
49383 50143: contig of 761 bp in length
50144 50243: gap of 100 bp
50244 50985: contig of 742 bp in length
50986 51085: gap of 100 bp
51086 51819: contig of 734 bp in length
51820 51919: gap of 100 bp
51920 52661: contig of 742 bp in length
52662 52761: gap of 100 bp
52762 53485: contig of 724 bp in length
53486 53585: gap of 100 bp
53586 54325: contig of 740 bp in length
54326 54425: gap of 100 bp
54426 55125: contig of 700 bp in length
55126 55225: gap of 100 bp
55226 55939: contig of 714 bp in length.

FEATURES
Source
1. .55939
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

Query Match 51.5%; Score 26.8; DB 2; Length 55939;
Best Local Similarity 73.9%; Pred. No. 5e+02;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 2 TGCAGTTCTCAGCGCTCTCATCGCTCTCGTCTCGCTCGCGCTCGC 47
Db 37378 TGGTGTCTCTTTCGCGCCCCCTCTCTCTCTCGTCTCGCTCGC 37423

RESULT 9
AC098566

LOCUS
DEFINITION
AC098566
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
Source
gene
CDS

AC098566 148301 bp DNA linear PLN 14-FEB-2002
Oryza sativa chromosome 10 clone OSJNBa0051J07, complete sequence.
AC098566
AC098566.3 GI:18390165
HTG.
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 148301)
Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H.-R., Rambo,T.,
Saski,C., Henry,D., Oates,R. and Simmons,J.
Rice Genomic Sequence
Unpublished
2 (bases 1 to 148301)
Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H.-R., Rambo,T.,
Saski,C., Henry,D., Oates,R. and Simmons,J.
Direct Submission
Submitted (24-OCT-2001) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
3 (bases 1 to 148301)
Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H.-R., Rambo,T.,
Saski,C., Henry,D., Oates,R. and Simmons,J.
Direct Submission
Submitted (29-JAN-2002) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
4 (bases 1 to 148301)
Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H., Rambo,T.,
Saski,C., Henry,D., Oates,R., Simmons,J., Thurmond,S.K. and Sun,S.
Direct Submission
Submitted (12-FEB-2002) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
5 (bases 1 to 148301)
Wing,R.A., Yu,Y., Soderlund,C., Chen,M., Kim,H., Rambo,T.,
Saski,C., Henry,D., Oates,R., Simmons,J., Thurmond,S.K. and Sun,S.
Direct Submission
Submitted (14-FEB-2002) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
On Jan 29, 2002 this sequence version replaced gi:18554329.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by more than
one plasmid subclone, and the assembly was confirmed by restriction
digest. The following are single sub-clone areas: 1616-1640,
13395-13416, 49634-49704, 67439-67512, and 67613-67642. There are
48 bases of TA Tandem Repeat at 110013-110061.

Location/Qualifiers
1. .148301
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4530"
/chromosome="10"
/clone="OSJNBa0051J07"
complement(266..2295)
/gene="OSJNBa0051J07.1"
/note="Contains similarity to wall-associated protein
kinase"
complement(join(266..315,745..930,1296..1551,1621..2295))
/gene="OSJNBa0051J07.1"
/codon_start=1
/product="Putative wall-associated protein kinase"
/protein_id="AAL77114.1"
/db_xref="GI:18652480"
/translation="MLMILPCRLLLRLVLATLLILAASTDQHVATLPITPLGCDKIC
GNISIPFPGTKQSCFLPGVNTDPTSPRLFLGNSRGNDRQNYQFEERYVLT
BEGMPHTLWTDDELFWELMSINLTGVARAYGVSDDCSINDTYHLVQKMTGLAGPF
LLISTRNALTAVGNWNEANLARSVRGSLKTCGRUGQPEFATNGSCLGCGCCQGET
QGIGSIADWNTSPDLYSDNFSKKYPKGVPLVIDFAIRDGFCPAGQAPANYACV
SSNSGVNTYNGDGYICNSGKGYDGNPIPNCHDIDECALRPLRVLYPCSRNG
ICMNRPGGYDPCPKRGMGDKAGTCSEKFPLOAKIVGDRVVKDKWDEVENTE"

ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; Oryza.
 1 (bases 1 to 306067)

REFERENCE
 AUTHORS The Rice Chromosome 10 Sequencing Consortium
 CONSRMT In-depth view of structure, activity, and evolution of rice
 TITLE chromosome 10
 JOURNAL Science 300: 1566-1569 (2003)
 REFERENCE
 AUTHORS Buell, C.R., Wing, R.A., McCombie, W.R., Messing, J. and Yuan, Q.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAY-2003) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

COMMENT This is the pseudomolecule for rice chromosome 10, which was
 constructed by resolving discrepancies between overlapping BACs,
 trimming the overlap regions, and linking the unique sequences to
 form a contiguous sequence. Genes in individual BAC clones were
 identified by a combination of several methods: Gene prediction
 programs, searches of the complete sequence against a peptide
 database and EST databases. Genes with similarity to other proteins
 are named after the database hits. Genes without significant
 peptide similarity but with EST similarity are named as unknown
 proteins. Genes without protein or EST similarity, that are
 predicted by more than two gene prediction programs over most of
 their length are annotated as hypothetical proteins. Genes
 encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>).

FEATURES
 source
 1. 306067
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /culivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="10"
 misc_feature
 <1..100654
 /note="Chromosome Sequence Derivation: nucleotide sequence
 in this region was derived from BAC clone OSJNBa0061H20
 (GB:AC113337)."
 <3288..>3647
 /locus_tag="OSJNBa0061H20.7"
 /note="hypothetical protein"
 <3288..>3647
 /locus_tag="OSJNBa0061H20.7"
 3288..3647
 /locus_tag="OSJNBa0061H20.7"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAP52356.1"
 /db_xref="GI:31430447"
 /translation="MGFYPAATTTLSTAOCVLVVGIGILLARATVROEHGQRLPLPS
 GSGRRQLRLSSNYPAGNQLRRQLRRDRARHYVWGHGSAHLGGGAIAIPDVLVG
 GLBERCIKPAQCPSAS"
 <8040..>11514
 /locus_tag="OSJNBa0061H20.8"
 /note="hypothetical protein"
 join(<8040..8132,8749..8836,9821..9901,10626..10672,
 11374..>11514)
 /locus_tag="OSJNBa0061H20.8"
 join(8040..8132,8749..8836,9821..9901,10626..110672,
 11374..11514)
 /locus_tag="OSJNBa0061H20.8"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAP52357.1"
 /db_xref="GI:31430448"
 /translation="MAPAGRRRLASKGSGNHALPSRSGRAGAMCAKGMGTWLF
 FSDAGDMETCGSRGDKGLLRQAQSCLLRRMHPLPGASVMRGCMQKNVISVGL
 QENYRHSINDVLWWSQKLDSDPVSNTAISLLQPSDEQAQVIM"
 complement(<21422..>28083)
 /locus_tag="OSJNBa0061H20.9"
 /note="Similar to RETROTRANSPOSABLE ELEMENT TF2 155 KDA

PROTEIN E value e 126 Identities 264 1333 (of subject);
 contains pfam domain: rvt Reverse transcriptase (RNA
 dependent D e value 4 1e 55; contains pfam domain: rve
 integrase core domain e value 6 1e 36; Co"
 complement(join(<21422..21824,22056..25325,25369..25920,
 26188..26993,27928..>28083))
 /locus_tag="OSJNBa0061H20.9"
 complement(join(21422..21824,22056..25325,25369..25920,
 26188..26993,27928..>28083))
 /locus_tag="OSJNBa0061H20.9"
 /codon_start=1
 /product="putative retroelement"
 /protein_id="AAP52358.1"
 /db_xref="GI:31430449"
 /translation="MSMRAGVLLLGAGQCLPVPGVTAAGVIGSVLLSIARMGKMLCH
 VFRLVTVMAAGSVSRRLDMITGFLEURMCVVLGVPHGVDTQARPIUPEQGGDAEPY
 AAEVTVAVILTSPERTSLAVTAGGDFPAACSAALLAIGTLHORVDELOHSPYR
 HPRRGARDYATFRDASDDATIVHLACVAYDAARIDFHMVRGMVNNKILE
 LRQENLQKDLDAEAQLKIAQGEVCRKRRVCRSOKITARKSTSRGLVROS
 LAWTCMETPRAEPAPMPOEGASGVSTEDALLTFRGPSSORRSVNTCDGNOPEG
 SNHNNGSPPPPPPPPPPDNAILTOILAQANMTAFHLHLQNLQNAAPPPP
 QHSLAEFLRIRPTFTSSNPVDALMDLHVAHGKKLDTVOCSDEKVIFAAHQJGPA
 SLWYHFQATQEQGPITWARPTAARFTHVPAGVALKKREFLESHKRRLSNMEMCG
 KMRRSRKSGMDPELSVRLVSGDYPDFQRLVDKLSIRLEAKHKELESKRRRLANFNQO
 GANQRYTYNYPGSSSQOQOQPSAPRQPFVVRVPOQOQSOQGTAPRPSTP
 TVQGGRRDAQOQRLCFNCFEGHFAKCPKRRQOQAPPRSNNGGKDVIRGRVN
 HVTAEVDLTPDVIUVGTFLLHSIPATILFDSGASHSFI SVFVGRNQLGVERLNPL
 ITTPGVMTAKYSPVPIEIQIPFSDLLDITKNDLVILGNMAQFQGVVDCAE
 RTVLYRGPEQVPVFPAPTSSVSELHQIGLSIPIVREFGDFPEELPGMPKREI
 EFRDILAPGTTPLYKPYRMAANELAEVKLELKEKGYIRPSTSPWGAIVIVEKK
 DTKRMCDYRLANNEVTINKYPLPRIDDLFDOLKGATVFSKIDRSQGHOLRIRED
 IPKATFTRVGLYEFTVMSFGLTNAPAFNNLNKVPMEYLDKFWVVIDDILYSQT
 EEDQHHLRLVLKLEHQLYAKLSKEFNLSEVKFLGHVISAQVADVPETVATD
 WROPKVTQIRSLFGLAYYRRIENFSIARPMTOQLLKEEFWSQCEKATOTLK
 EKLVSFVLIPDTRKDFVMVYCDAAQGLCVLMQDGHVAVASRQLRPHREGNPTHD
 LELAAVHALKIWRHLYINCCEIYTDHLSKYIFTOSDLNLRORRWLELTKDYDVG
 LHYHGRKANVADALSRSKNTLNVRGIPLELNOQMEALNSI VGRGFALAEAKPTL
 DDYIRQAQNDPMRGLLKNMKQKATGFTDEHGTLMGNRCVCPDNRELKQLIQE
 AHSPYSIHGSGTMYLDLKKYVMSMKREIAEFVLCDCVQRKAEHQRPGLQP
 LQFPEWKDGIKMDFITLPTKQGGYSIWWVDRLTKVARFIPVKTTYGKNKLAELY
 FARIVSLHGPKXIVSDRGSQFTSHFMKLOEELGTRLNFTAYHPTQDGTGTERLNOI
 LEDMLRACVLDLFGTKSLPYAEFSYNNYSQASIQMAPYEALFQTKLAPFVGPFP
 RIIRARGEVAYQLLEPASLGNVHDVHVSQLKCLRPSEQADSEQIEVRDLTVVER
 PVKILDMERRTRNRYIRFCVKQVSNHAEAEATEREDELKAAHPDLFASSESRGRD
 SV"
 complement(<34426..>35324)
 /locus_tag="OSJNBa0061H20.10"
 /note="hypothetical protein"
 complement(join(<34426..34543,34635..34835,35161..>35324))
 /locus_tag="OSJNBa0061H20.10"
 complement(join(34426..34543,34635..34835,35161..35324))
 /locus_tag="OSJNBa0061H20.10"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAP52359.1"
 /db_xref="GI:31430450"
 /translation="MEAVARHLGHVLSAPTNEHDCSVTLPMNSGATAPRSLYVP
 PPUEBETLVALKLSQKASTFPDTHAAGRHPLHWDSPATOLEDLNPLEEYIGIQ
 AAVGSFNSFSNSKEAVRGRGLATYTLSSRRVPASGDAEDFVPCDGHLMQVGKEQBERG
 "
 complement(<36243..>36653)
 /locus_tag="OSJNBa0061H20.11"
 /note="hypothetical protein"
 complement(<36243..>36653)
 /locus_tag="OSJNBa0061H20.11"
 complement(36243..36653)
 /locus_tag="OSJNBa0061H20.11"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAP52360.1"
 /db_xref="GI:31430451"
 /translation="MDESDLOWRQGMVAVGLPPAPVAVSAIVSFYLTWQMDYE
 EQLRRRTQAGLWVLLVGLAVALVLLGSHALVDAGRVAVPVSNRWGGSGYSSGDDGG

are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

- NOTE: This is a 'working draft' sequence.
- This sequence will be replaced
- by the finished sequence as soon as it is available and
- the accession number will be preserved.

FEATURES	source	Location/Qualifiers
BASE COUNT	37722 a 28659 c 29003 g 39973 t	
ORIGIN		1. .135357 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="genomic DNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /chromosome="2" /clone="P0017C12"

Query Match	50.4%	Score	25.2	DB 2	Length	135357			
Best Local Similarity	72.3%	Pred. No.	7e-02						
Matches	34	Conservative	0	Mismatches	13	Indels	0	Gaps	0

Oy 2 TGCAGTTCTCTCAGGCTCTCAICGCTCTCGTCGCTGCCGGGCTCGCC 48
||||| ||| ||||| ||| ||| ||||| |
D6 35193 TGCAGGTCTCTCAGGCCCTCCAGGCCCAAGTGC GCCACGGCTCTGCC 35239
||||| ||| ||||| ||| ||| ||||| |

RESULT 13				
AC140331				
LOCUS	AC140331	230357 bp	DNA	linear
DEFINITION	Mus musculus chromosome UNK clone RP23-406J20, WORKING DRAFT SEQUENCE, 7 unordered pieces.			HTG 23-FEB-2003

AC140331
AC140331.1 GI:28475582
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
SOURCE

ORGANISM	Mus musculus	Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		

REFERENCE	1 (bases 1 to 230357)
AUTHORS	McPherson, J.D. and Waterston, R.H.
TITLE	The sequence of Mus musculus clone

REFERENCE
 2 (bases 1 to 230357)
AUTHORS
 McPherson, J.D. and Waterston, R.H.
TITLE
 Direct Submission
JOURNAL
 Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

```

COMMENT
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
-----
Project Information
Center project name: M BA0406J20
-----
parkway, St. Louis, MO 63106, USA
-----

```

```

Center project name: M_BA006020

----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer Et; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 229568 bases at least Q40
Consensus quality: 230223 bases at least Q30
Consensus quality: 230705 bases at least Q20
Insert size: 196000; agarose-fp
Insert size: 244718; sum-of-contigs
Quality coverage: 17.67 in Q20 bases; agarose-fp
Quality coverage: 11.05 in Q20 bases; sum-of-contigs

```

* NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces

- * is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

• 1 1255: contig of 1255 bp in length
 • 1356 1355: gap of unknown length
 • 1356 21952: contig of 20597 bp in length
 • 21953 22053: gap of unknown length
 • 22053 52473: contig of 30421 bp in length
 • 52474 52573: gap of unknown length
 • 52574 95111: contig of 42538 bp in length
 • 95112 95211: gap of unknown length
 • 95212 135051: contig of 39840 bp in length
 • 135052 135151: gap of unknown length
 • 135152 170910: contig of 35759 bp in length
 • 170911 171010: gap of unknown length
 • 171011 230357: contig of 59347 bp in length

```

LOCATIONS/Qualifiers
1. 230357
source
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP21-406J20"

```

misc_feature	1. .1255	/note="assembly_name:Contig28"	1356. .21952	/note="assembly_name:Contig37"	22053. .52473	/note="assembly_name:Contig33"	52574. .95111	/note="assembly_name:Contig34"	95212. .135051	/note="assembly_name:Contig35"	135152. .170910	/note="assembly_name:Contig36"	17011. .230357	/note="assembly_name:Contig37"	230358. .260544	/note="assembly_name:Contig38"	260545. .290792	/note="assembly_name:Contig39"	290793. .321040	/note="assembly_name:Contig40"	321041. .351288	/note="assembly_name:Contig41"	351289. .381536	/note="assembly_name:Contig42"	381537. .411784	/note="assembly_name:Contig43"	411785. .442032	/note="assembly_name:Contig44"	442033. .472280	/note="assembly_name:Contig45"	472281. .502528	/note="assembly_name:Contig46"	502529. .532776	/note="assembly_name:Contig47"	532777. .563024	/note="assembly_name:Contig48"	563025. .593272	/note="assembly_name:Contig49"	593273. .623520	/note="assembly_name:Contig50"	623521. .653768	/note="assembly_name:Contig51"	653769. .684016	/note="assembly_name:Contig52"	684017. .714264	/note="assembly_name:Contig53"	714265. .744512	/note="assembly_name:Contig54"	744513. .774760	/note="assembly_name:Contig55"	774761. .805008	/note="assembly_name:Contig56"	805009. .835256	/note="assembly_name:Contig57"	835257. .865504	/note="assembly_name:Contig58"	865505. .895752	/note="assembly_name:Contig59"	895753. .926000	/note="assembly_name:Contig60"	926001. .956248	/note="assembly_name:Contig61"	956249. .986496	/note="assembly_name:Contig62"	986497. .1016744	/note="assembly_name:Contig63"	1016745. .1046992	/note="assembly_name:Contig64"	1046993. .1077240	/note="assembly_name:Contig65"	1077241. .1107488	/note="assembly_name:Contig66"	1107489. .1137736	/note="assembly_name:Contig67"	1137737. .1167984	/note="assembly_name:Contig68"	1167985. .1198232	/note="assembly_name:Contig69"	1198233. .1228480	/note="assembly_name:Contig70"	1228481. .1258728	/note="assembly_name:Contig71"	1258729. .1288976	/note="assembly_name:Contig72"	1288977. .1319224	/note="assembly_name:Contig73"	1319225. .1349472	/note="assembly_name:Contig74"	1349473. .1379720	/note="assembly_name:Contig75"	1379721. .1409968	/note="assembly_name:Contig76"	1409969. .1440216	/note="assembly_name:Contig77"	1440217. .1470464	/note="assembly_name:Contig78"	1470465. .1500712	/note="assembly_name:Contig79"	1500713. .1530960	/note="assembly_name:Contig80"	1530961. .1561208	/note="assembly_name:Contig81"	1561209. .1591456	/note="assembly_name:Contig82"	1591457. .1621704	/note="assembly_name:Contig83"	1621705. .1651952	/note="assembly_name:Contig84"	1651953. .1682200	/note="assembly_name:Contig85"	1682201. .1712448	/note="assembly_name:Contig86"	1712449. .1742696	/note="assembly_name:Contig87"	1742697. .1772944	/note="assembly_name:Contig88"	1772945. .1803192	/note="assembly_name:Contig89"	1803193. .1833440	/note="assembly_name:Contig90"	1833441. .1863688	/note="assembly_name:Contig91"	1863689. .1893936	/note="assembly_name:Contig92"	1893937. .1924184	/note="assembly_name:Contig93"	1924185. .1954432	/note="assembly_name:Contig94"	1954433. .1984680	/note="assembly_name:Contig95"	1984681. .2014928	/note="assembly_name:Contig96"	2014929. .2045176	/note="assembly_name:Contig97"	2045177. .2075424	/note="assembly_name:Contig98"	2075425. .2105672	/note="assembly_name:Contig99"	2105673. .2135920	/note="assembly_name:Contig100"	2135921. .2166168	/note="assembly_name:Contig101"	2166169. .2196416	/note="assembly_name:Contig102"	2196417. .2226664	/note="assembly_name:Contig103"	2226665. .2256912	/note="assembly_name:Contig104"	2256913. .2287160	/note="assembly_name:Contig105"	2287161. .2317408	/note="assembly_name:Contig106"	2317409. .2347656	/note="assembly_name:Contig107"	2347657. .2377904	/note="assembly_name:Contig108"	2377905. .2408152	/note="assembly_name:Contig109"	2408153. .2438400	/note="assembly_name:Contig110"	2438401. .2468648	/note="assembly_name:Contig111"	2468649. .2498896	/note="assembly_name:Contig112"	2498897. .2529144	/note="assembly_name:Contig113"	2529145. .2559392	/note="assembly_name:Contig114"	2559393. .2589640	/note="assembly_name:Contig115"	2589641. .2619888	/note="assembly_name:Contig116"	2619889. .2650136	/note="assembly_name:Contig117"	2650137. .2680384	/note="assembly_name:Contig118"	2680385. .2710632	/note="assembly_name:Contig119"	2710633. .2740880	/note="assembly_name:Contig120"	2740881. .2771128	/note="assembly_name:Contig121"	2771129. .2801376	/note="assembly_name:Contig122"	2801377. .2831624	/note="assembly_name:Contig123"	2831625. .2861872	/note="assembly_name:Contig124"	2861873. .2892120	/note="assembly_name:Contig125"	2892121. .2922368	/note="assembly_name:Contig126"	2922369. .2952616	/note="assembly_name:Contig127"	2952617. .2982864	/note="assembly_name:Contig128"	2982865. .3013112	/note="assembly_name:Contig129"	3013113. .3043360	/note="assembly_name:Contig130"	3043361. .3073608	/note="assembly_name:Contig131"	3073609. .3103856	/note="assembly_name:Contig132"	3103857. .3134104	/note="assembly_name:Contig133"	3134105. .3164352	/note="assembly_name:Contig134"	3164353. .3194600	/note="assembly_name:Contig135"	3194601. .3224848	/note="assembly_name:Contig136"	3224849. .3255096	/note="assembly_name:Contig137"	3255097. .3285344	/note="assembly_name:Contig138"	3285345. .3315592	/note="assembly_name:Contig139"	3315593. .3345840	/note="assembly_name:Contig140"	3345841. .3376088	/note="assembly_name:Contig141"	3376089. .3406336	/note="assembly_name:Contig142"	3406337. .3436584	/note="assembly_name:Contig143"	3436585. .3466832	/note="assembly_name:Contig144"	3466833. .3497080	/note="assembly_name:Contig145"	3497081. .3527328	/note="assembly_name:Contig146"	3527329. .3557576	/note="assembly_name:Contig147"	3557577. .3587824
--------------	----------	--------------------------------	--------------	--------------------------------	---------------	--------------------------------	---------------	--------------------------------	----------------	--------------------------------	-----------------	--------------------------------	----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	-----------------	--------------------------------	------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	--------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------	---------------------------------	-------------------

Query Match	50.4%	Score 26.2	DB 2	Length 230357
Best Local Similarity	79.5%	Pred. No. 6.7e+03		
Matches 31	Conservative	0	Mismatches 8	Indels 0
				Gaps 0

Qy 1 ATGCAGTTCTCTCAGGCTCTCATCGCTCTCGTCGTGCC 39
 ||||| ||||| | | | | | |||||
Db 107371 ATGCAGTGTCTCACCCACTGAGCGATCTCCCGTGCC 107409

RESULT 14

LEISERTHRE

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

REFERENCE

[illegible]

ATTI

JOURNAL

JOURNAL
MEDLINEMEDLINE
PUBMED

COMMENT

•

FEATURES
source

Location/Qualifiers
1. .2595
/organism="Leishmania donovani"
/isolate="MHOM/BR/82/BA-2"
/sub_species="chagasi"
/db_xref="taxon:5661"
/dev_stage="promastigote"
/germline
891. .2111
/function="serine/threonine phosphatase"
/standard_name="PP2C"
/codon_start=1
/evidence=experimental
/product="protein phosphatase-2C"
/protein_id="AAA02864.1"
/db_xref="GI:293068"
/translation="MGIPLPKPVMTQLQERYGNAIFRCGSCNVNGYRETMEADHLTYL
TDSMGFFGVDFGHVNDOSQYLERWASAEIKESI PMTDERMKELALRIDQWMDSGR
EGSGTGFVVALKEGKVKHLQGVNDSRVVACIDGVCVPLTEDHKPNNEGERQRIEN
CAGRVNRRVDGSLAVSRAGDREYKLGSSQLEKQVIALADVQHKFTFDSNDFVLL
CCDGVFGNFPNEEVVAYVQQLETCNDLAEVAGRVCEEAIERGSRDNI SCMIVQFKD
GSQYAAEPHTTVPGFSPAPRNSGFRKAYESMADKNTTGVALLERRYDTLKAAEALT
PEETELSLQFENGPEAKLTGAEROKWFSNYFQKLC EAASNGSPSDMERLQSLQQQAGI
PLSTLLSLMGEQTQ"

CDS

BASE COUNT 555 a 797 c 687 g 556 t
ORIGIN

Query Match 50.0%; Score 26; DB 3; Length 2595;
Best Local Similarity 76.2%; Pred. No. 1.1e+03;
Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 8 TCTCTACGCTCTCATCGCTCTCTCGCTGCGCGCCCTCGCCA 49
Db 493 TCGCTCTCGCTCTCATCGATCCAGTCGCCCCCTCCCTCCCA 534

RESULT 15
AC073431

LOCUS AC073431 117846 bp DNA linear HTG 12-JUL-2001
DEFINITION Homo sapiens chromosome 8 clone GSI-146A9, WORKING DRAFT SEQUENCE,
6 unordered pieces.

AC073431
VERSION AC073431.10 GI:14702002
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 117846)
McCombie, W.R.

AUTHORS Human Genomic Sequence, Chromosome 8
TITLE Unpublished
JOURNAL 2 (bases 1 to 117846)
REFERENCE McCombie, W.R.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA

COMMENT On Jul 12, 2001 this sequence version replaced gi:14423582.
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory

Center code: CSHL
Web site: <http://www.cshl.org/genseq>
Contact: mcombie@cshl.org
----- Project Information
Center project name: GSI-146A9
Center clone name: GSI-146A9

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 81715: contig of 81715 bp in length
* 81716 82745: gap of unknown length
* 82746 98894: contig of 16149 bp in length
* 98895 99924: gap of unknown length
* 106720 107749: contig of 6796 bp in length
* 107749 112805: contig of 5057 bp in length
* 112806 113834: gap of unknown length
* 113835 115554: contig of 1720 bp in length
* 115555 116583: gap of unknown length
* 116584 117846: contig of 1263 bp in length.

FEATURES
source

1. .117846
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/clones="GSI-146A9"
/clone_lib="Genome Systems 1"
31403 a 25767 c 26372 g_28909 t 5395 others

BASE COUNT 31403 a 25767 c 26372 g_28909 t 5395 others
ORIGIN

Query Match 50.0%; Score 26; DB 2; Length 117846;
Best Local Similarity 70.0%; Pred. No. 8.1e+02;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 ATGCAGTCTCTCAGCGCTCTCATCGCTCTCGTCGCTCGCGCCCTCGCCAG 50
Db 1416 ATGGAGTTCTCTGAACCTCTGCTGCTCTGAGGCTGCTGACTCTCGAG 1465

RESULT 16
AP005635

LOCUS AP005635 149082 bp DNA linear HTG 08-AUG-2002
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 9 clone P0646C08,
*** SEQUENCING IN PROGRESS ***.

AC005635
VERSION AP005635.1 GI:22138859
KEYWORDS HTG; HTGS PHASE2.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1
Sasaki, T., Matsumoto, T. and Katayose, Y.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, PAC
clone:P0646C08

AUTHORS Published Only in Database (2002)
TITLE 2 (bases 1 to 149082)
JOURNAL Sasaki, T., Matsumoto, T. and Katayose, Y.
AUTHORS Direct Submission
TITLE Submitted (07-AUG-2002) Takuji Sasaki, National Institute of
Agricultural Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>,
Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1. .149082

RESULT 18	TITLE
AC084081/C	JOURNAL
LOCUS	COMMENT
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	


```

repeat_region      /rpt family="LTR36"
16283..16415
/rpt family="L2"
repeat_region      complement(16758..17011)
/rpt family="LTR37a"
17012..17044
/rpt family="T-rich"
repeat_region      complement(17045..17191)
/rpt family="LTR37a"
17545..17838
/rpt family="AluSx"
repeat_region      complement(18260..18562)
/rpt family="AluJo"
repeat_region      complement(18764..18965)
/rpt family="L1MD3"
18966..19348
/rpt family="MSTA"
repeat_region      complement(19349..19654)
/rpt family="AluSp"
19655..19695
/rpt family="MSTA"
repeat_region      complement(19696..19835)
/rpt family="L1MD3"
repeat_region      complement(19829..20084)
/rpt family="L1MC4a"
repeat_region      complement(20255..20878)

Query Match      50.0%; Score 26; DB 9; Length 211291;
Best Local Similarity 70.0%; Pred. No. 7.7e+02;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGAGTTCTCTACGCTCTCATCGCTCTCTGCTCGCTCGGCTCGGCGAG 50
|||||
Db 24135 ATGAGTTCTCTGAACCTCTGCTGCTGCTGAGGCTCGCTGACTCTGAG 24184
|||||

RESULT 21
OCU18102/c      6000 bp mRNA linear MAM 08-AUG-2001
LOCUS      Oryctolagus cuniculus mRNA for titin.
DEFINITION      Y18102
ACCESSION      Y18102.1 GI:3928488
VERSION      titin; titin gene.
KEYWORDS      Oryctolagus cuniculus (rabbit)
SOURCE      Oryctolagus cuniculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE      1
AUTHORS      Gregorio,C.C., Trombitas,T., Kolmerer,B., Stier,G., Granzier,H.,
Kunke,K., Suzuki,K., Obermayr,F., Herrmann,B., Sorimachi,H. and
Labeit,S.
TITLE      The NH2 terminus of titin spans the Z-disc: its interaction with a
novel 19-kD ligand (T-cap) is required for sarcomeric integrity
J. Cell Biol. 143 (4), 1013-1027 (1998)
JOURNAL      99034591
MEDLINE      9817758
PUBMED      9817758
REFERENCE      2 (bases 1 to 6000)
AUTHORS      Labeit,S.
TITLE      Direct Submission
JOURNAL      Submitted (16-SEP-1998) S. Labeit, EMBL Heidelberg, Meyerhofstr. 1,
69117 Heidelberg, FRG
COMMENT      Related sequences U28657, D83390, X90568.
FEATURES      Location/Qualifiers
source      1..6000
            /organism="Oryctolagus cuniculus"
            /mol_type="mRNA"
            /db_xref="taxon:9986"
            /chromosome="2"
            /map="2q24"
            /tissue_type="soleus skeletal muscle"
            /dev_stage="adult"
            1..6000
            /gene="titin"
gene

```

```

CDS
1..>6000
/gene="titin"
/codon_start=1
/product="titin"
/protein_id="CAA77028.1"
/db_xref="GI:3928489"
/db_xref="SPTREMBL:O97791"
/translation="MTTQALFTFQLOPSVVVLEGSAAATFEAHVSGFVPEVSVWRDQO
VNSTSLPGVQISFSDGRKALTI PAVTKANSGRYSLRATNGSOATSAELLVLAETA
PINFVORLQSMVROGSQVRLQVRVTGIPTPVVKFYRDGAEIIOSSLOFQISQEGDLVS
LRIAEAYPEDSGTYSVYNATSVGRATSTAGQLVQGEVPEVPAKTKTITVSTQIAKQO
TRIEKIEAHGFDARSINATVMVVDGATGQOLPHKTPPIIPKPKSRSPPTSPSIAAKQO
LARQOSPSPIRHSPSPVRHVPAPTPSPRVSVPAGRISTSPISVKSPLLVRKTQTIT
MATGPEVPPKQOEGVVASSTAEAMRETTMTSSTEIRREWEGRYGLQBOVITIGAA
GAAAVSTSAFAGLGSVFATGKQADKSAAVTAAVVAADVMDVARYREPIVSAVEQ
TAQRTTTTAVHIQPAQEQVRKEAEKIAVSKVVAADAKAEQELKSRREITTTKQEOQ
HI THEOLRKETAKAFVPKVVISAAKAKOEOTRITGEITTKQEOQKQIOTETMMKETRT
VVPKVIATPKVKEQDLVSRREGISTKREQVOITOEKMRKEAEKTLSTAVATAKA
TEQETVLRREGNATQEQIQVTHGKVGUGKAEAVTAAVVAADVQARVREPREPLGE
DSYAQQTTLIEYGKEHISATKVAEHPKRPASEPHVVPKAVPVVQAPSETHIKTTEQ
VGMHISQIKKTTDVTSERLVHVDRKPRTPASPHFTVSKI SVPTKEHGYEASIAGSAIE
TLQKELSATSSAOKITKSVKAPTVPKPAEAVRAEPTSPQPFPAATPDYTKSQAGIEV
KKEGVSVITGSAVREERLEVLHGREAQVEMARVPAEIPVTPPTLVSLGKNMTVIE
GESVTLECHISGVPSTVWYRDYQIESIQITFQSGIARLMI REAFADRGRT
CSAVNAGVTSTCYLAVQVSEFEKETTVAELKTTTEKRFVESRDVMDTTTIEEQ
AMPGEAPAFPFTIKPVVQKLVGGSVVFCQIGGNPKPHVYWKSGVPLTTGYRYKVS
YNKOTGECKLVI SMTFADDAGETYIVIRNKHGETSASALLEAEYESLMSKQOEMLY
QTQMTAFVQEPKVGIEAPGVSYSEYEKEQALIRKMAKDTVMVTRTFVEDQEFHI
SSPEERLKEIEYRIIKTTLVEELLEEDGEEKMAVDISESAIESGFELRIKNYRILEG
MGVTFHCKMNGYPLKLIAYKDGKRIKPGERYOMDFLODGRASIRIPVLPEDGELIIP
RGTYRAKEIAAPALEPLHLYRGOEEDGLDYKQKQKPFKKKLSLRLKRFQPAH
PECELTPIGDPMTVENVLHDGKPLEANRLRMINEFGYCSLDYGVAVSRDSGVLTCRA
TNKIGTDHVSATLIVKDESLVESQIPGKQGLQRIEELERMAHEGALCTVTDQRE
KQKPDILVLPVLEGEFRVLEGTFRVTPGYPPQKLNMYNLQGLIRKSKRFRVADGHI
YLDIVCKSYDGTGVKTAENPEGVIEHKVKLEIQOERDFSVLRRAPEKPEFPHIE
SGKLQFEVQKVRPVDTSKTEVVKLRAERI THEKVSSESESLRSRPFKRRTESGYE
AITAVELKSRKKDSEYELLKTKDELHHTKELTAEEKAL"
BASE COUNT      1859 a 1402 c 1502 g 1236 t 1 others
ORIGIN
Query Match      49.2%; Score 25.6; DB 4; Length 6000;
Best Local Similarity 77.5%; Pred. No. 1.4e+03;
Matches 31; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2 TGCAGTTCTCTCAGGCTCTCATGCTCTCTGCTCGCTGCGG 41
|||||
Db 2633 TGCAGTACTTCCAGGCGCTCTCGGCACGCGCTGCGCG 2594
|||||

RESULT 22
AE008035/c      10951 bp DNA linear BCT 18-DEC-2001
LOCUS      Agrobacterium tumefaciens str. C58 circular chromosome, section 93
of 254 of the complete sequence.
ACCESSION      AE008035
VERSION      AE008035.1
KEYWORDS      GI:15156068
SOURCE      Agrobacterium tumefaciens str. C58 (Cereon)
ORGANISM      Agrobacterium tumefaciens str. C58 (Cereon)
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
REFERENCE      1 (bases 1 to 10951)
AUTHORS      Hinkle,G., Slater,S.C. and Goodner,B.
TITLE      Complete Genome Sequence of Agrobacterium tumefaciens C58
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 10951)

```


	/note="hypothetical protein RP464 - Rickettsia prowazekii"									
	/codon_start=1									
	/transl_table=11									
	/product="AGR_C 1943p"									
	/protein_id="AAK86861.1"									
	/db_xref="GI:15156077"									
	/translation="MSSAFADRLIRVSGTGAEEFLNNLTADINLPEGETRASAL									
	LTPGKLIPLDFELIWRDGRDYLVETGAEEQDALLRRLLTWYKLRARVELKAEVIGVGF									
	WGNVITAGVRDGRFACAGVGVVPGASGEAAAEALRVHGHGIAESGRDYALQDA									
	FPDVLMDVNDGVSKGCFVGVQVWVRMKHRTARRRVTVTSADGTLPASGTETAN									
	GKPVGALCTVYGNRALAIVTRDVRADALAAGTPLLADNVAVSVALPAWSGLSFPAADP									
	AARAED"									
gene	complement (8769..9536)									
	/gene="AGR_C 1946"									
CDS	complement (8769..9536)									
	/gene="AGR_C 1946"									
	/note="hypothetical protein"									
	/codon_start=1									
	/transl_table=11									
	/product="AGR_C 1946p"									
	/protein_id="AAK86862.1"									
	/db_xref="GI:15156078"									
	/translation="MCSLSKGLRSSALLSILLATAGPIGLPEKAFALSCLKPAQKG									
	ESEPAQNGEQEQAQPAEDVEGDSGIPLPDPLVDRSAGKAOQSAKDDTTPEMS									
	VWIEHDSKAPVPRKLRQOIIDAASGDIKLRPFIVAGONOFRIDGSEDPIAL									
	KSYSGDPDGLVLAIIIDLLSTGYAHIDAGTPTDEAYVFPYPAGKPLNTLTAPEKVELL									
	RIITAGDLMQEYGNYSFYRIGISPDKWKKFFTAGD"									
	complement (9551..9808)									
gene	/gene="AGR_C 1947"									
	complement (9551..9808)									
CDS	/gene="AGR_C 1947"									
	/note="hypothetical protein"									
	/codon_start=1									
	/transl_table=11									
	/product="AGR_C 1947p"									
	/protein_id="AAK86863.1"									
	/db_xref="GI:15156079"									
Query Match	49.2%; Score 25.6; DB 1; Length 10951;									
Best Local Similarity	70.8%; Pred. No. 1.3e+03;									
Matches	34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;									
Oy	2 TGCAGTCTCTCACGCTCTCATCGCTCTCGTCGCGCGCGCTCGCCA 49									
	-									
Db	6308 TGCAGACGGCTCGATTCTCGCCGCTCGCGCGCTCGCCA 6261									
RESULT 23										
AE009069/c										
LOCUS	11764 bp DNA linear BCT 20-DEC-2001									
DEFINITION	Agrobacterium tumefaciens strain C58 circular chromosome, section									
	95 of 256 of the complete sequence.									
ACCESSION	AE009069 AE008688									
VERSION	AE009069.1 GI:17739437									
KEYWORDS										
SOURCE	Agrobacterium tumefaciens str. C58 (U. Washington)									
ORGANISM	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;									
	Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.									
REFERENCE	1 (bases 1 to 11764)									
AUTHORS	Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,									
	Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,									
	Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,									
	Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,									
	McClelland, E., Palmeri, A., Raymond, C., Rouse, G.,									
	Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,									
	Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,									
	Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,									
	Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.									
	and Neeter, E.W.									
TITLE	The genome of the natural genetic engineer Agrobacterium									
JOURNAL	tumefaciens C58									
	Science 294 (5550), 2317-2323 (2001)									

MEDLINE	21608550	
PUBMED	11743193	
REFERENCE	2 (bases 1 to 11764)	
AUTHORS	Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,	
	Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,	
	Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,	
	Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,	
	McClelland, E., Palmeri, A., Raymond, C., Rouse, G.,	
	Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,	
	Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,	
	Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,	
	Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.	
	and Neeter, E.W.	
Direct Submission		
Submitted (27-SEP-2001)	Department of Microbiology, University of	
Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA		
98195-7242, USA		
FEATURES	Location/Qualifiers	
source	1..11764	
	/organism="Agrobacterium tumefaciens str. C58 (U.	
	Washington)"	
	/mol_type="genomic DNA"	
	/strain="C58"	
	/db_xref="taxon:180835"	
gene	177..1220	
	/gene="Atul046"	
CDS	177..1220	
	/note="identified by sequence similarity; putative; ORF	
	located using Blastx/Glimmer"	
	/codon_start=1	
	/transl_table=11	
	/product="ABC transporter, substrate binding protein"	
	/protein_id="AAL42059.1"	
	/db_xref="GI:17739438"	
	/translation="MILTRLLSRTLAIGAAGVAGSLPATGAYETPVKFTLDWKF	
	EGPAGELLALDKGYFKAEGLDVTDIDSGNSVEAIPRVATGAYOMGFGDINSVMKFLD	
	EDFSQKIKAVMYTERPTFAVGRKSGVDGPKSLGKGLGAPPPGAFQWPAFKQ	
	VAGLDISKIKIESIGFPVPEMLAKGDVAVGFAFVILNKKOGIADDDISTILMA	
	EHLNLYGNNAVLVNTDFAEKNPFAVKGLKALAKGFADSVKTPPEGVAALVLRNETLD	
	SAIELERLNMGNSMNIKTPYVVENGMGVDPARLAASLDTLKVSMLGKGNVKAQGVFD	
	AAYLPKKEERMLP"	
	1471..2259	
gene	/gene="Atul047"	
	/note="identified by sequence similarity; putative; ORF	
	located using Blastx/Glimmer"	
	/codon_start=1	
	/transl_table=11	
	/product="ABC transporter, nucleotide binding/ATPase	
	protein"	
	/protein_id="AAL42060.1"	
	/db_xref="GI:17739439"	
	/translation="MSHLVIDSDMYGSGAGTLVAGLNLAVDKCEFAAVGPGSGC	
	GKSTLMKLVTLGIHPQRGNVIVAGRQVTEPVSIVGMAFQNPMTMLPWRITLNLIPLE	
	IVEKHRRRLRAHKASYAKAELELLEIVLGLGFGSKYQWLSGSGMOORANLCRLIHQP	
	ELMLDPEFGALDAFTREELWCVRIDHAAQRTIILVTHDLREATFLADKIFVMSAR	
	PGRVLAERHVPFARPROLDIMYDKGFNDMVQELHGEIAQARVAA"	
	2256..3074	
gene	/gene="Atul048"	
	/note="identified by sequence similarity; putative; ORF	
	located using Blastx/Glimmer"	
	/codon_start=1	
	/transl_table=11	
	/product="ABC transporter, membrane spanning protein"	
	/protein_id="AAL42061.1"	
	/db_xref="GI:17739440"	
	/translation="MSAIMESSVAAEPRKPLIERVNVNKAAPMLYTLALFALWELLVY	
	ALKMPPTILPSIRVGOVIVQYWSPIWNSLQTLYTTTLGFAIIVAAGLAGLGLFGWS	
	KTIYAGLYPLMIGFNAIPKVALPILVIFGIGTGPVAVLTAFLISFFPIVNVATGLA	

TIEPTEDVILRALGAKKNDIMLVKGIIPRSMPIYFPGSLKIAITLAFVGSVVSETVASNY
 GLGNMMSAQSOFNVLPIYFAGLLMLAVEGIAIYAMVMAWLEKRMITGWHRSTMGO"
 complement (3126. .3809)
 /gene="Atu1049"
 CDS
 complement (3126. .3809)
 /gene="Atu1049"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAL42062.1"
 /db_xref="GI:17739441"
 /translation="MQKTLTRVTCALAGSLMVAAAAATAGCDKPAAHSHNHGHDS
 QMTEAEKIYKGFEDAGKRPRLTDEGDMQSVPIYLLNGSLDSVMAEKAAGHDKTA
 AEYRAYEIGYKDTDNRIIVAPGSVTFYKEGKPGKTVYSDGOELITYKKGNGRVYI
 FKKTAGYDAEAPQFIQDSHTITAPDKAGHYHLVNGNDRKALMDVEINMPTYYPFLMDGA
 DIVEEMAAH"
 complement (3987. .4808)
 /gene="Atu1050"
 CDS
 complement (3987. .4808)
 /gene="Atu1050"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAL42063.1"
 /db_xref="GI:17739442"
 /translation="MEQQTADILVASOTAVRQASALAVQYSFVIGALLLLIIGWLAA
 AFLRFWAFQGLSRIGRQFATILGAFAGIRVYVILVIMVLQGFQVQTASILAALCA
 LIGALQGLTQNIIAAGIMLLVLRPFVGEYIETSTVGTCTIEIGLFATELKTSDG
 LVRLAPNSTLANVPIYNSRPSRHLSLTKNEDMAAADMVTVRSESRILL
 PAPVTFVDSVTADSATVKLYRWVRSNDYFVTRDVTKAMRLAFDERKAENVQAQA"
 complement (4886. .5404)
 /gene="Atu1051"
 CDS
 complement (4886. .5404)
 /gene="Atu1051"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAL42064.1"
 /db_xref="GI:17739443"
 /translation="WTCIVTSPSLRIAEAMVKKAREMTVLIAKEQSFHRPAPVVAAR
 HLTAMNDLIAFKGTGDIAPDPAHLVLIIDFAGEMDSAPLLIHCWGISRSPAAVAI
 AALSLYPDDESTELAQRLRTVSPVYVNSRIIAGDRLLGREGLVKAIKAIGRGADT
 DGNVPFVULPLAG"
 complement (5401. .6024)
 /gene="Atu1052"
 CDS
 complement (5401. .6024)
 /gene="Atu1052"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAL42065.1"
 /db_xref="GI:17739444"
 /translation="WAPAKSPRAWQMLSGRRLLDLPDVEIADIAHGLARVARW
 NGOTRGDHAFTVAOHCILVETIFCRMCPGATPEMHWALHDAPEVIGDMISPPKSV
 VGGGKYTKVEKLEAAVHLRGLPSPHASERLKDRIFKADTVAAFFTEATELAGFSTAEAQ
 KFGGLPGKITHDMFDIPLRGLSTEAQRLFIARFEAIETLRLVTRTGGAV"
 complement (6027. .6875)
 /gene="gcvt"
 /note="synonym: Atu1053"
 CDS
 complement (6027. .6875)
 /gene="gcvt"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11

/product="glycine cleavage system T protein, aminomethyltransferase"
 /protein_id="AAL42066.1"
 /db_xref="GI:17739445"
 /translation="MSSAFADLRRLRVSGTGAEEFNLLITADINLPEGETRASAL
 LTPQKILFDLIRQDRDYLVBETGAEDQALLRRLTYKLRPAVEKKAETVEGIGV
 WGNSTVTEAGVDRGRFAGAKGVDRFVPGASGSAEAAEYALRVHEGSGRDYALODA
 FPHDVLNDVNGVSFKKGKGFVGQEVVSRMKHRCGTARRRVTVTSADGTLPASGTETAN
 KPVKGALETVYGNRALAIVRTDRVADALAAGTPLLADNVAVSVALPAWSGLSPFPAADP
 AARAD"
 complement(6998..7765)
 /gene="Atul054"
 complement(6998..7765)
 /note="Atul054"
 /notes="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAL42067.1"
 /db_xref="GI:17739446"
 /translation="MCSLQGLLRSSALLSILLALTAGPIGLPEKAFALSELKPAQKQ
 ESFPQANEQPOQAQPADEVEGSDGIPDPDVRSGAKAQSASAKPDTTPMS
 VLEHDIKAPAEVRKLRQOIIDAAASGDIKLRPFIVAGNQFRIDGSDSDPIAAL
 KSYSGDDGLEVLAIIDLLSTGYAHIDAGTDEAYVFPYFAGKPLNLTITAPEKVELL
 RIITAGDLADMOEYGNYSFVRIGISPDGKWKFFTAGD"
 complement(7780..8037)
 /gene="Atul055"
 complement(7780..8037)
 /gene="Atul055"

Query Match 49.2%; Score 25.6; DB 1; Length 11764;
 Best Local Similarity 70.8%; Pred. No. 1.3e+03;
 Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 2 TGCAGTTCCTCAGCTCTCATCGCTCTCTCGTCGCCGCTCGCCA 49
 ||||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
 Db 4537 TGCAGACGGCTCGATTCTCGCGCCCTCGCGCTCGCGCTCGCCA 4490

RESULT 24
 AP005454
 LOCUS
 DEFINITION
 AP005454
 PAC clone:P0455F03, complete sequence.
 AP005454
 AP005454.2 GI:29824104
 HTG.
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 1
 Sasaki,T., Matsumoto,T. and Katayose,Y.
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
 clone:P0455F03
 Published Only in Database (2002)
 2 (bases 1 to 106732)
 Sasaki,T., Matsumoto,T. and Katayose,Y.
 Direct Submission
 Submitted (20-JUN-2002) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program, Kanmoudai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
 Tel:81-298-38-7441, Fax:81-298-38-7468
 On Apr 11, 2003 this sequence version replaced gi:21624006.
 The orientation of the sequence is from SP6 to T7 of the PAC clone.
 FEATURES
 source
 1..106732
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"

LOCUS DEFINITION	AP003569	164978 bp	DNA	linear	HTG 21-MAR-2002
ACCESSION	Oryza sativa (japonica cultivar-group)	chromosome 6 clone P0425F05	*** SEQUENCING IN PROGRESS ***		
VERSION	AP003569				
KEYWORDS	HTG; HTGS_PHASE2				
SOURCE ORGANISM	Oryza sativa (japonica cultivar-group)				
REFERENCE	Oryza sativa (japonica cultivar-group)				
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.				
TITLE	1				
JOURNAL	Sasaki, T., Matsumoto, T. and Yamamoto, K.				
AUTHORS	Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC				
TITLE	clone:P0425F05				
JOURNAL	2 (bases 1 to 164978)				
AUTHORS	Sasaki, T., Matsumoto, T. and Yamamoto, K.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-MAY-2001) Takuji Sasaki, National Institute of				
COMMENT	Agrobiological Resources, Rice Genome Research Program; Kannondai				
	2-1-2, Tsukuba, Ibaraki 305-8602, Japan				
	(E-mail: tsasaki@affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,				
	Tel: 81-298-38-7441, Fax: 81-298-38-7468)				
	NOTE: It currently consists of 1 contigs. Gaps between the contigs				
	are represented as runs of N. The order of the pieces is believed				
	to be correct as given, however the sizes of the gaps between them				
	are based on estimates that have provided by the submitter. This				
	sequence will be replaced by the finished sequence as soon as it is				
	available and the accession number will be preserved.				
	* NOTE: This is a 'working draft' sequence.				
	* This sequence will be replaced				
	* by the finished sequence as soon as it is available and				
	* the accession number will be preserved.				
FEATURES	Location/Qualifiers				
source	1..164978				
	/organism="Oryza sativa (japonica cultivar-group)"				
	/mol_type="genomic DNA"				
	/cultivar="Nipponbare"				
	/db_xref="taxon:39947"				
	/chromosome="6"				
BASE COUNT	47386 a 34709 c 34874 g 47659 t	350 others			
ORIGIN					
Query Match	49.2%;	Score 25.6;	DB 2;	Length 164978;	
Best Local Similarity	77.5%;	Pred. No. 1e+03;	9;	Indels	0; Gaps 0;
Matches	31;	Conservative	0;	Mismatches	0;
QY	9	CTCTACCGCTCTCATCGCTCTCGCTCGCGCGCGCGCGCC	48		
Db	17406	CGCTCATGCTCTCATCGCTCTCGCTCGCGCGCGCGCGCC	17367		
RESULT 27					
AC114896	Oryza sativa (japonica cultivar-group)	chromosome 3 clone	HTG 08-MAY-2002		
LOCUS	OSJNBa0091B22, *** SEQUENCING IN PROGRESS ***	5 ordered pieces.			
DEFINITION	AC114896				
VERSION	AC114896.3	GI:20503086			
KEYWORDS	HTG; HTGS_PHASE2				
SOURCE ORGANISM	Oryza sativa (japonica cultivar-group)				
REFERENCE	Oryza sativa (japonica cultivar-group)				
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.				
	1 (bases 1 to 118444)				
	Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Tallon, L., Gansberger, K.,				
	Kim, M., Overton II, L., Bera, J., Tsitrin, T., Krol, M., Jarrahi, B.,				
	Jin, S., Koo, H., Zismann, V., Hsiao, J., Blunt, S., Vanaken, S.,				
	Utterbach, T., Feldblyum, T., Yang, Q., Haas, B., Suh, B., Peterson, J.,				
	Quackenbush, J., White, O., Salzberg, S. and Fraser, C.				

AUTHORS TITLE JOURNAL

Buell,R.
Direct Submission
Submitted (08-MAY-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On May 8, 2002 this sequence version replaced gi:20146765.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

COMMENT

1 21285: contig of 21285 bp in length
* 21286 21385: gap of unknown length
* 21386 58264: contig of 36879 bp in length
* 58265 59364: gap of unknown length
* 59365 71533: contig of 13169 bp in length
* 71534 71633: gap of unknown length
* 71634 73850: contig of 2217 bp in length
* 73851 82956: gap of unknown length
* 82957 83066: contig of 9016 bp in length
* 83067 118460: contig of 35394 bp in length
* 118461 118560: gap of unknown length
* 118561 144077: contig of 25517 bp in length.

FEATURES

source
1. 144077
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="3"
/clone="OSJNBA0036E17"
/note="japonica cultivar-group"

BASE COUNT 41309 a 31223 c 30860 g 40084 t 601 others
ORIGIN

Query Match 48.8%; Score 25.4; DB 2; Length 144077;
Best Local Similarity 68.6%; Pred. No. 1.2e+03;
Matches 35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 TGCAGTCTCTCAGCGTCTCATCGCTCTCGTGGCGGCTCGCCAGTG 52
Db 135221 TGCAGTCTCTCAGCGTCTCTGTGCTCAGCTGCGAGCGTCTCGCTCTG 135171

RESULT 30 CNS08C87/c

LOCUS CNS08C87 152359 bp DNA linear HTG 07-MAR-2003
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 12 clone
OSJNBA0029K06, *** SEQUENCING IN PROGRESS ***.

ACCESSION

AL731758

VERSION

AL731758.3 GI:28892654

KEYWORDS

HTG; HTGS PHASE2; HTGS ACTIVEPIN.

SOURCE

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 152359)

Choisine,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P.,
Seguens,B., Pelletier,E., Scarpelli,C., Salanoubat,M.,
Weissenbach,J. and Quetier,F.

Oryza sativa chromosome 12 sequencing

Unpublished

2 (bases 1 to 152359)

Genoscope.

Direct Submission

Submitted (06-MAR-2003) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT

On Mar 9, 2003 this sequence version replaced gi:28460535.
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: <http://www.genoscope.cns.fr/>
Contact: SegRef@genoscope.cns.fr

The following sequence is oriented from the T7 to the SP6 end.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc... even if efforts are made to eliminate these
contaminating sequences.

The nucleotide sequence of this BAC clone was generated by
combining Syngenta and Genoscope sequencing data.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 152359: contig of 152359 bp in length.

FEATURES

source
1. 152359
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="12"
/clone="OSJNBA0029K06"
/clone_lib="OSJNBA"

BASE COUNT 43054 a 34046 c 33464 g 41741 t 54 others
ORIGIN

Query Match 48.8%; Score 25.4; DB 2; Length 152359;
Best Local Similarity 68.6%; Pred. No. 1.2e+03;
Matches 35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 TGCAGTCTCTCAGCGTCTCATCGCTCTCGTGGCGGCTCGCCAGTG 52
Db 110057 TGCAGCGCGCTCGCGTCTTGTGCTCTCGCCGAGCGTCTTGTCTCTTG 110007

RESULT 31 CNS08CAF/c

LOCUS CNS08CAF 168864 bp DNA linear HTG 07-MAR-2003
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 12 clone
OSJNBA0083D13, *** SEQUENCING IN PROGRESS ***.

ACCESSION

AL831802

VERSION

AL831802.4 GI:28892660

KEYWORDS

HTG; HTGS PHASE2; HTGS ACTIVEPIN.

SOURCE

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 168864)

Choisine,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P.,
Seguens,B., Pelletier,E., Scarpelli,C., Salanoubat,M.,
Weissenbach,J. and Quetier,F.

Oryza sativa chromosome 12 sequencing

Unpublished

2 (bases 1 to 168864)

Genoscope.

Direct Submission

Submitted (06-MAR-2003) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)


```

identity in 342 aa overlap). Contains Pfam matches to
entry PF00532 Peripla_BP_like. Periplasmic binding
proteins and Laci family and to entry PF00356 laciI,
Bacterial regulatory proteins, laci family. Contains
probable helix-turn-helix motif at aa 9-30 (Score 2307,
+7.04 SD)
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB45587.1"
/db_xref="GI:5139569"
/translations="MTETGPRPTLEAVERAGVSRTSVRVNGHGVDRDALAERVRR
AVEELGVNPHARSVLVTRHOAVAVVAEPETRVFADPVFAOQLAGISKELTAHDNO
LVLLLEGEDHARVGRYLACGHVDGALVESLHDPGLGVSAGVPTVFGGRPDND
DGRDDVVYDSDNRGNARSVAHHLAGLGRTRIAHTIGPLDQTSAAADLAGFRDVRGA
GVLVARDMLTSGGGERAMRELLDPCGLDAVFAANDLTAAAGALRVLRGRRRVDDV
AVVGFDDMLPVAEQTPPLTTVTRQDIEGMRLMARLLRLGLDRAADATDFAARTPT
AAPGGVVLPTTLVHRSTA"
1291..1374
/gene="SC06713"
/notes="Pfam match to entry PF00356 laciI, Bacterial
regulatory proteins, laci family, score 39.70, E-value
9.4e-10"
1465..2202
/gene="SC06713"
/notes="Pfam match to entry PF00532 Peripla_BP_like,
periplasmic binding proteins and Laci family., score
95.60, E-value 9.5e-25"
/complement(2088..2111)
/notes="repeat 7; 23 bp imperfect inverted repeat"
/gene="SC06714"
/notes="synonym: SC4C6.24c"
/complement(2440..3231)
/gene="SC06714"
/notes="SC4C6.24c, possible hydroxylase, len: 263 aa;
similar to TR:Q55078 (EMBL:U50973) Streptomyces sp. strain
C5 daunomycin C-14 hydroxylase (275 aa), fasta scores:
opt: 422 z-score: 482.3 E(): 1.5e-19, 38.5% identity in
273 aa overlap. Similar to TR:Q50527 (EMBL:AL009204)
S.coelicolor hypothetical protein (267 aa) (37.9% identity
in 253 aa overlap) and to (EMBL:AL049863), SC5H1.09c.
S.coelicolor possible hydroxylase (265 aa) (35.7% identity
in 258 aa overlap). Also similar to hypothetical proteins
from Mycobacterium tuberculosis and to Mycobacterium
tuberculosis 27.3 kDa MAB HBT7 reactive antigen
(EMBL:AJ007737) (260 aa) (36.7% identity in 264 aa
overlap)"
/codon_start=1
/transl_table=11
/product="putative hydroxylase"
/protein_id="CAB45588.1"
/db_xref="GI:5139570"
/translations="MLTTRVTVGAPNWLVDGTSYGLGLFQWRFRSAGPEAG
GYAFELDRIVAGMTTEEQPPSWTVYFOAPDARAAQAAGHGVLVGPMDVM
DQGTMAILADRALGFCIMQPGORAGLDVTGECALCWELHTADTAAAAAYRAVLG
LETSGVSPFGSVCVNPAGEGDAMFGGLVPLAEDPADTDADAGLPPYFAVDDAOTA
VARTELGGTVRPATDIEGVRVARLDYPGARFAVLRPAPRQG"
/complement(3239..3244)
/notes="possible RBS"
3425..3670
/gene="SC06715"
/notes="synonym: SC4C6.25"
3425..3670
/gene="SC06715"
/notes="SC4C6.25, probable transcriptional regulator, len:
81 aa, similar to many transcriptional regulators (some
putative) from Streptomyces e.g. TR:Q53963 (EMBL:X62287),
WhiB, S.coelicolor transcriptional regulator essential for
sporulation (87 aa), fasta scores; opt: 336 z-score: 452.7
E(): 5.7e-18, 61.1% identity in 72 aa overlap. Also
similar to e.g. TR:Q69649 (EMBL:AL02212) Mycobacterium
tuberculosis putative regulatory protein (100 aa) (40.0%

```

```

identity in 75 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB45589.1"
/db_xref="GI:5139571"
/translations="MHTETTAAPDLNQBEALCAQTGGDFFFPFPGSSVRDAKRICAL
CPIRSTCLEFALSNDERFGVMGLSEKRLALRTTP"
4003..4197
/gene="SC06716"
/notes="synonym: SC4C6.26"
4003..4197
/gene="SC06716"
/notes="SC4C6.26, hypothetical protein, len: 64 aa; unknown
function, probable CDS suggested by positional base
preference, GC frame plot and amino acid composition"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB45590.1"
/db_xref="GI:5139572"
/translations="MCRAGGVAALLLTEAAAEPLAVTGALGECEDRPAPWRELPRAP
RRHPAPGVRGAAPGEITAE"
complement(4242..5228)
/gene="SC06717"
/notes="synonym: SC4C6.27c"
complement(4242..5228)
/gene="SC06717"
/notes="SC4C6.27c, probable acyl-(acyl-carrier protein)
desaturase, len: 328 aa; similar to many e.g.
SW:STAD.LINUS (EMBL:X70962) Linum usitatissimum
stearoyl-ACP desaturase (396 aa), fasta scores; opt: 385
z-score: 461.5 E(): 2.2e-18, 30.2% identity in 311 aa
overlap. The N-terminal 100 aa are divergent. Highest
Query Match 48.8% Score 25.4; DB 1; Length 292200;
Best Local Similarity 68.8%; Pred No. 1.1e+03;
Matches 35; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
Qy 1 ATCGAGTTCTCACGCTCTCATCGCTCTCGTCGTCGCGGCGCTCGCCAGT 51
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279401 ATCGGTACGCCAACGCGCGCGCTCGTCGCTCGCTCGCTGT 279351

RESULT 33
AE005738/c
LOCUS
DEFINITION
AE005738 Caulobacter crescentus CB15 section 64 of 359 of the complete
genome.
ACCESSION
AE005738 AE005673
VERSION
AE005738.1 GI:13421828
KEYWORDS
SOURCE
ORGANISM
Caulobacter crescentus CB15
Caulobacter crescentus CB15
Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
Caulobacteraceae; Caulobacter.
REFERENCE
1 (bases 1 to 10037)
Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R.,
Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H.,
Shetty,J., Berry,K., Ufferback,T., Tran,K., Wolf,A., Vamathevan,J.,
Ermoiaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
Fraser,C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
21173698
MEDLINE
11259647
PUBMED
REFERENCE
2 (bases 1 to 10037)
Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,

```

gene	4459. .7353 /gene="CC0623" 4459. .7353 /gene="CC0623" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAK22608.1" /db_xref="GI:13421832"
CDS	/translation="MSLHKIEIFEDEICAHLAANGWLYDCAAGDRHRTALYVEDLV AWQDSQPEAMDGLSRANGAAQVLADRVKSLDERGLVDVLRNGVELVGARGRIQM AQPKPAMGNPEIVRRYEANRLVARQVYKAVNGENCLDLVLYNLGLVPVATLAEKSD TQSVEDAINQYRFDREPRPKGEEPLAPFGGALVHFVAVSNSVAMTTKLGDRDTFF LPNKGDHGKGNPLNPNHRTAYLWEEIWARDSWLEILGRYMWTKRDAKRTISVVF PRYHOLDATKLAQVLAEGAGGKFLIHSAGSGKTSIAWSAHLISELHDAADKNLKF SSVIVSDRNVIDTQLEALDFORTAGVATIKISGSGSKSQGLAEALAAAGKVVICT IQTFPALEAVREQAATQGAFAFIADENHSSQTGEAGKLNKMLSDAEIAELADGGE IDFETQLTANMAGRAADKSTIYVATATPKAKTLQLPGRDPADGLPAPFHVYSRQA IEEGFILDVKNVTPYSLAFKLANDGOEWDERVEDRTALKGIMRWVRLHPVYIAQKV QIVVEHYRENVQPLLDCKAMVVLGSRQEAVRWSVAINAYIRKMGYPLGLTVAFSGE VNDELGAVSETSAALNPGNLNGRDIRDAFKAENPHFLLVANKFQTFQDPLCCGMVVD RRLAGIQAQVTLRLNRAHPKQDITTYILDFCNSAEDILTAFTYETFALEAATDNL IYDLRAKLAAGHYDDHEIERVVRVELDPKAKQGDILAALEPVADELKAFKAAQKAW VTKAODDERAAEALDHNMALDFKGDMAAFQRLYSFLSQIFDYGNTEIEKRSIFFR RLPLLEFRGERPVDLSGLRLTHALKTKGQODLLYGTGEAPKLSGMAEYVSGGLHE KEKARLDEIIAKLNDLPEGEITEENTCSTSEVLQPGCGFSGGLASSGFDI" complement(7350. .8330) /gene="CC0624" complement(7350. .8330) /gene="CC0624" /note="identified by match to protein family HMM" /codon_start=1 /transl_table=11 /product="ISCC3, transposase OrfB" /protein_id="AAK22609.1" /db_xref="GI:13421833"
gene	/translation="MRYPACEKIEIIRLVEQSTLPVRRTLAQIGVPRATFFRWVDLYO TGGPEALDRPAPSRVVMNRIPEDEVRAAILMALEHSELSPRELAFTDSQGFVSE ASVYRLKSHDLITSPAFIVMKAAKAFTKTAPNOMQWDTFTFKIIGMWYLSIT LDDFSRYVIAWKLTCTTKWAGDVTDTLELASAGCQVSVHRKPLRLSDNGSSYISGE LADWLKGQEMKHVYGPHTQTKIERWHQTLKNRILLENYFLPGLDEAQIGAFIEH YNHARYHLENTVPADAYFGRADILRRRASVKRQTLRHLRRLQHRKIAA" complement(8327. .8704) /gene="CC0625" complement(8327. .8704) /gene="CC0625" /note="identified by match to protein family HMM" /codon_start=1 /transl_table=11 /product="ISCC3, transposase OrfA" /protein_id="AAK22610.1" /db_xref="GI:13421834"
CDS	/translation="MRPKPLAAKATPGERVVKDIRRATRRHFSAEKIRIILDGLRGE DSIAELCRKEGIAQSLYVYVWSKEFMEAGKRRLAGDTARAATTDVEKDLRRESTALKEV VARQALEIRLLKKSMIADGGGE" complement(8745. .9119) /gene="CC0626" complement(8745. .9119) /gene="CC0626" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK22611.1" /db_xref="GI:13421835"
gene	/translation="MADRKPGDSVMIMNETPPDITFVIGDHPDPSLMDLPWINRAEVAP AGEGRDDEIRLVKLFRRGENNELVSSALPPLPDRPHDLVAYNVEGAPQVFWHTLRFV EAYRIDPDCPPSAPMQRIEGSV" complement(9136. .9471) /gene="CC0627" complement(9136. .9471) /gene="CC0627"
CDS	

gene	3405. .4448 /gene="CC0622" 3405. .4448 /gene="CC0622" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK22607.1" /db_xref="GI:13421831"
CDS	/translation="MTRLIVIRAGRHCERELEAITQGRLLPGFGEVADLTGLKDREAI LRVLVAITPEGATLURNFAAQMNFVNTIGVGLDVVLPRKVTSGVAIGRVTPGYRHD PAAPAIATRAVEWLEAVNEAPFAQDLRHSFGAFLTCEVKRNEALARVEAVLTKGD PGPLLGKVPYVKAPEVDAEADVELDVEDIARQOIIALIKSAFAGHDLAYLVAEIL RIEYVTVKSPGPGGVDTILAAAGRGLGEDRICVQVKSADAPANQEVVILKMGVQ AAKATGLLIVSINGVNGPAQLIDSEFFSVRLQWPELLTALTALFASYQLSDETRAKIP LKQIWPVAVGDA"

TITLE	Laub, M.T., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
JOURNAL	Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J.,
FEATURES	Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J.,
source	Ermlaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and Fraser, C.M. Direct Submission Submitted (31-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA Location/Qualifiers 1. .10037 /organism="Caulobacter crescentus CB15" /mol_type="genomic DNA" /strain="CB15" /db_xref="taxon:190650" 224. .2059 /gene="CC0620" 224. .2059 /gene="CC0620" /note="identified by match to protein family HMM" /codon_start=1 /transl_table=11 /product="type I restriction-modification system, M subunit, putative" /protein_id="AAK22605.1" /db_xref="GI:13421829" /translation="MNHQSLGAFISWADGLVGRYKQSDYGVKILPFTVLRRLDCVLE PTKAAVLAEHEKRAAGVDPPEFLRRVSGAGFYNTKMDLTVLGDADNVAANLYAYL QATSPARDIIFERPEFHAQIERLAKGLVYVAEKFTRIDLPEAVNHOMHGVFEEL IRKFAELNSETAGEHFTPREVIRLMVELIIVEDDAALSEGVVRTIYDPTAGTGMLS VADERLLQNPAGKLSYNGELNPESAICADMLIKQDPDNIIVFNTLSDDGHAA KFDYMLSNPFFGVEWKKVIVRAEASQQYNGRFGPLPRVSDGSLFLHLLSKMR PAVDGSRKFLGSLPTGGAGSGEISRRHLLSDLVEAIVALTDFMTFNTGIAT YVMIVSNKGAARGLQLIDASGFMRKRSKSKRKEGDEDDIAAITRLFGAFVEA ELASVDAEGKPDVIVKVEAGSAPPAPEGVKVLAPLSKILPNSAFGYRTITVERPL VDEAGKPLVGKGNKCKPOADSARDTENVLSEDEVTFAREVLPHPADAMIDADK TKGTGEIPFNRHFVFPFPRDLAQIDADLRATVDQIKAMIAELAA" 2056. .3408 /gene="CC0621" 2056. .3408 /gene="CC0621" /note="identified by match to protein family HMM" /codon_start=1 /transl_table=11 /product="type I restriction-modification system, S subunit" /protein_id="AAK22606.1" /db_xref="GI:13421830" /translation="MSFPAYESYKESGVLEWLRVPSHWRPRLKHLVIMRSGGTPSK REDYMGGEIPWASAKDLKVDLTIDTQDLTAELDEGAQQLPANAVVILVRGMMLAR TFPVLRLSRMTINQDLKLIANGVDPNYLAWSLRASEVETLCRLDEAGHGTKALRM DAWSTWELPAPSLAEQQAIAAFIDRETAKIDALVEAQERLIALLKEKRAVISHAVTK GLDPSAQMDSGVWELQMPAHWEVVPKKNLADSIKAGPFGSALTMDMYSSAGRVYVG QEOVIGDPRIGDYVYTSDRYNELSOYRVVEGDLLVSCVCTCKIAIFPOGAPGLIN PLRIFRPNQVDPYLCVLLRSANVSFEQSYLSRGSTMDVINIGLGEIVPVPFMQ EQISIAGLAEVQEQFDLSAASEAAITLLQERRAALISAATVGTGKIDVRLGVLTAEAA" 3405. .4448 /gene="CC0622" 3405. .4448 /gene="CC0622" /note="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK22607.1" /db_xref="GI:13421831"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="AAK22612.1"

/db_xref="GI:13421816"

/translation="MGLGMSNAAPSARYPTSKVPVADLILRVCADLDRWTVRCPLHGVARQLFGPELAARKLADVLYKLPERGAFKCKAQYGCNGVPAISEISVDAMDVGQ LQNVACWSR"

complement(9893)..9969)

/gene="tRNA-Arg-4"

complement(9893)..9969)

/gene="tRNA-Arg-4"

/product="tRNA-Arg"

1874 a 3295 c 3142 g 1726 t

BASE COUNT

ORIGIN

Query Match 48.5% Score 25.2; DB 1; Length 10037;

Best Local Similarity 90.0% Pred. No. 1.7e+03;

Matches 27; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 23 TCGCTCTCGTCGTCGCCGCGCTCGCCACTG 52

Db 1311 TCGCTCTCGCGCTCGCGCGCCCGCCGCGTG 1282

RESULT 34

AP003047/c

LOCUS

DEFINITION

Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC clone:P0666G04.

AP003047 BA000010

AP003047.2 GI:13122417

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

ORIGIN

[illegible]

```

/codon_start=1
/product="hypothetical protein"
/protein_id="CAB95347.1"
/db_xref="GI:9366585"
/translations="MRSRYMRKGNESVTVAVSFTREDAFFLSCMRSLGIGVAPYLRC
RSPQDFPLKFLSTLAETPENRIVAAPCTTWTVDVDDHDDVEGLMPLQSCQAC
SPHVKSHFOLLVLSYFAIQACRTTSTSMKHWQRLSPATFKSDVEDAAATFLOI
LSENSVPPNELMNCRYTQTHSCRLTKDRMLKNLKEGPVLAUMLPLVDLMISQNVAE
GNVALKRCDAQRSLMRSLNPLKCAKOHLSVCDDDAAVWUULETVVAVKEYIPLNL"
2218
/notes="Majority of WCS, TPB15A20, TP7E4 and TPB12C6; reads
call T, Two TP20A12 reads call C,
type 1 (single nucleotide substitutions)"
complement(2222..5611)
/genes="TB927.1.540"
complement(2222..5611)
/genes="TB927.1.540"
/EC number="2.7.7.6"
/codon_start=1
/product="DNA-directed RNA polymerase III, probable"
/protein_id="CAB95348.1"
/db_xref="GI:9366586"
/translations="MTPRARGDEGYATRKALLPALINMKGLLNHHVASFDDLINVE
LGRILNESNREIKSSIDPDFSIYSGIRVCRPREIVGKQLPKVPSPHECRTRDLTY
RGDIVDVEYTSRDSKRICETETNIKIGTIPMLKSSCNLYGKGREELTAMRECPD
PGCYFIIKGVEKVLQEOEQKNRVIIIEADDNGDIVAHVQSKTHYSISKCVTRFKGR
IVMHRSPKEDIPIVVVLKALHESDQOIAHQHIGVSPAFOSVLPACFEHAASLUNVKTQ
DDALHFTIGERRKETFMEDETQQRHVQSKADSAAEFTLANVLICHIREGEVQRDMNFR
HKALYVCFMVRVEASVSSLDLDERDFYGNKRFEITTLGLMLLFEDLLKQYRNVVKT
AMQLLASKDDTTFPNFLQJMESKEVINGMRVALSSGRWELKRFNMSRQGITQVLS
RLSVISICITMTTLASSFEKSRKISGRSLQPSQWGMVCPDTPEGESCLGVKNFAL
SOYTLMDSDSFVRAAYVNLGVEVDCI SPNEFLKYTVFLNGTLMGHRYPSRLRGV
RTLRSRGLPHVSIALNDRQKSVQICDGRVRLYIVVRDKPATIVSOHLNLSG
KCSINDFLAGLVEFDINEANDCLIAVYPDIEQYTHLEIEPLSLVAGVIGI PPP
HHNOSARNTFQSAMGQALGTIAVNOYIRMDTVMLGAYPQRLCTRKAANLTNYEIL
GAGINAMCVMSYSGYDIEDAQVYNRSLDRGYRCVLRKHEVELERLPNGSYDII
PADSNVGRSALNSDGIASKGAVVQFDILVNKYTPVPSMEPRNPVLVYKPOPAPVD
HVVISPPGSDSRMDVEOKI KVTTRVREPGDKFSSRHGQKGVGLITTDGVNMEFN
EOGICPDMPNNPFPSPMTVQKLLLELVSSKVALCKGFGDTAFGDKAEDLCEELI
RHGYNHGKDI FVSGVTGEMLRAYVFPPIYIYORLKHMTDKMHRATGRPSMLTQOP
TEGRSRSGURGEMERDCMVGASNLNLERLLVSSDLFTDIDICCCGPMGYNGVCS
YCNKNKTVSRVNPYPFKLLQLQNGIGSTRLSLDPMSQ"
3257
/notes="Majority of WCS, TP3D18, TP7E4 and TPB15A20; reads
call T, Two TP20A12 reads call C, One WCS read calls C
type 1 (single nucleotide substitutions)"
3343
/notes="Majority of WCS, TPB15A20, TPB12C6, TP3D18; and TP7E4
reads call C, One WCS read calls T
type 1 (single nucleotide substitutions)"
3734
/notes="Majority of WCS, TPB15A20, TP3D18 and TPB12C6; reads
call A, Three TP20A12 reads call G
type 1 (single nucleotide substitutions)"
5345
/notes="Majority of WCS, TPB15A20, TPB12C6, and TP3D18 reads
call A, One WCS read calls C
type 1 (single nucleotide substitutions)"
complement(5981..6283)
/genes="TB927.1.550"
complement(5981..6283)
/genes="TB927.1.550"
/notes="unlikely gene"
/codon_start=1
/product="hypothetical protein, unlikely"
/protein_id="CAB95349.1"
/db_xref="GI:9366587"
/translations="MDSFCAICLTFTCIDERGSARPCLSIIWVKHFWFHEGWADDEE
TPRFELHIRFDATAFVQMAVMGDFNLLYVGRASSGWGGEKKKTETLMLHYVS"
complement(6507..7835)
/genes="TB927.1.560"
complement(6507..7835)
/genes="TB927.1.560"

```

```

/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAB95350.1"
/db_xref="GI:9366588"
/translations="MLFVASSKSHLINLDTKTVSRVYTPTVERRGLVYMPARVVM
SNORSCASIFSPAMOQPVLCITPOATVATASSTDCGAFLLVGTAEGLNVWSTLSGE
LLHVRSHRRITTEIATISDQSLIATASEDSVCKTWLLAALVARGVSAPAPRIVPNGH
LSLVNACSLDSSQSVVATSSDRSKRIFDGTQQQVTVLGDVLTSVRSPGDEMLL
IGGETSLFFVRLYSFSEKQCLTAGLIRCVNDVVTCTPFTDGHKGAI VFIHEDYTRP
DYVLDTNGVILMMNIRTATVVEKAVDDIAGGLLSVYVPSDATOPSSAAVPPVLIQ
KHPLDPLGVDFIVVSAGKESAPISYQEGGSRASRRKRKHQOQTMEDTCVSMKRTENV
VATELEGARESTGKPTSAERFSRLREKQDELEFLKNLKEKLKLRAS"
7583
/notes="Majority of genomic shotgun, TP7E4, TP3D18, TPB12C6
and TPB15A20 reads call A, One genomic shotgun read calls
G
type 1 (single nucleotide substitutions)"
7896
/notes="Majority of WCS, TP3D18, TP7E4, TP3D18 and TPB12C6
reads call G, One WCS read calls A
type 1 (single nucleotide substitutions)"
complement(8396..9223)
/genes="TB927.1.570"
complement(8396..9223)
/genes="TB927.1.570"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAB95351.1"
/db_xref="GI:9366589"
/translations="MDARVHRSKRGCGRVLKRASSPEVVAEVTGCRPGDPFVLPRLR
SPNGSGCHCPHSCDCLYVSKELERCYLSQLSVQVTCCKRELCAKSVLPFPFVAR
LLRVNATDDTFYDUGCNGSVLFHFVALATGASCVGVEINEHNAKVAKEAWTHLRPV
FEKRRKGLDVSIVCGDFCKVQKODNYFSSSCVVMANLLMPRVFNHLYSERLSRPI
GSRVLCMEDLYPHSRVAAARDPDAFEKFEVMDYRWQEDSVESVSPASGPFVLYIKRS"
complement(10259..11809)
/genes="TB927.1.580"
complement(10259..11809)
/genes="TB927.1.580"
/codon_start=1
/product="phosphate-repressible phosphate permease,
possible"
/protein_id="CAB95352.1"
/db_xref="GI:9366590"
/translations="MWHPLYLWIIIVGIVSFLTCGCGVMNDLANSFGTTVYGSRVNLW
QIVILASICEFVGAVSGSEVTSTGGIANPMTFAPEVLYMYGMCAUSATFIWLL
FATMSLPSVSTHSIAGIIFALVYVGFAGFAKKIDFPVYTVGVAPITASFISF
MFAKVAASALYALLRLVLRPANVRNALPALIVGTVFFLEFVFLFKGAKARLHW
GPAKASAAALIGLGAASI SAACIPIILRRRVRLITERAERAEETGMNTAPEISGDAG
AISENAAGVCAAVEGPDVTANRIVPPSSPTSDPTAKENPANNASGLTTPGVVDEAL
KEDVOIYDERVEYVRYLQVTFACASFAHGANDVNAIAPFSNMYSIYINQOVVEEN
DVPMLVLVGGAGLVGLATLGVGIMRLGERITKTPSRGFSAEALSVALVVSUCSAP
GIPVSTHCITGAVVAISIMDCGFRKVRMMVMVGKYLGIWFTLLITAAISALLPAQGI
YAPSLTQ"
12152
/notes="Majority of WCS, TP3D18 and TP7E4; reads call C, Two
WCS reads call T,
type 1 (single nucleotide substitutions)"
13074
/notes="Majority of genomic shotgun, TPB15A20, TP3D18, and
TP7E4 reads call A, Two TP20A12 reads call C
type 1 (single nucleotide substitutions)"
complement(13430..13846)
/genes="TB927.1.590"
complement(13430..13846)
/genes="TB927.1.590"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD53050.1"
/db_xref="GI:25991393"
/translations="MNAFHGLCDHQRKSLRSVLLFPFISRQLIILALSVPYGTVLI
FPFPPSPFFFPFFFLCFICLSCLFSNLINLHTRTAPAPSAWIPTLWHSVARR
VALSGVEGCRVTDVRECLQNIMLEKVGSSSSSR"
13700..13709

```

variation

variation

gene

CDS

gene

CDS

variation

variation

gene

CDS

variation

```

/note="Majority of WCS,TP3D18,TP7B4,TPB15A20 and;TP16A7
reads call 10 G, One WCS and TP20A12 read call 12G

Query Match      48.5%; Score 25.2; DB 3; Length 281000;
Best Local Similarity 71.7%; Pred. No. 1.3e+03;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy  4  CAGTTCTCTACGCTCTCATCGCTCTCGTCGCTCGCGCCCTCGCCA 49
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   64495  CATTCTCTCTCGCTCTCATCACGCTCATCTCGCTCTCTCGTCA 64540

RESULT 36
PSU96179      723 bp  DNA  linear  BCT 06-NOV-2001
LOCUS      Pseudomonas syringae pv. syringae Hma (hrma) gene, partial cds;
DEFINITION  and unknown gene.
ACCESSION  U96179
VERSION    U96179.1 GI:2149939
KEYWORDS   Pseudomonas syringae pv. syringae
SOURCE     Pseudomonas syringae pv. syringae
ORGANISM   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
REFERENCE   1 (bases 1 to 723)
AUTHORS     Alfano,J.R., Klm,H.S., Delaney,T.P. and Collmer,A.
TITLE       Evidence that the Pseudomonas syringae pv. syringae hrp-linked hrma
            gene encodes an Avr-like protein that acts in an hrp-dependent
            manner within tobacco cells
JOURNAL     Mol. Plant Microbe Interact. 10 (5), 580-588 (1997)
MEDLINE     97348579
PUBMED     9204563
REFERENCE   2 (bases 1 to 723)
AUTHORS     Alfano,J.R. and Collmer,A.
TITLE       Direct Submission
JOURNAL     Submitted (02-APR-1997) Plant Pathology, Cornell University, 334
            Plant Science Bldg., Ithaca, NY 14853, USA
FEATURES    Location/Qualifiers
             1. .723
                /organism="Pseudomonas syringae pv. syringae"
                /mol_type="genomic DNA"
                /strain="61"
                /db_xref="taxon:321"
                /note="pathovar: syringae"
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AA858695.1"
                /db_xref="GI:2149940"
                /translations="MIDKAFALTLLRDDTHORLLLTGLLEPHEDLPLORLLAGALNP
                LVNAGPGIGHDEQSLGYHAYQSIPREKVSVMLEKLEIAGLVEMKWCWEART"
                /note="ORF1"
                /db_xref="taxon:321"
                /note="pathovar: syringae"
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AA858695.1"
                /db_xref="GI:2149941"
                /translations="MNPPIHA"
                /translation="MNPPIHA"
BASE COUNT  170 a 202 c 173 t
ORIGIN

Query Match      48.1%; Score 25; DB 1; Length 723;
Best Local Similarity 69.4%; Pred. No. 2.4e+03;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy  4  CAGTTCTCTACGCTCTCATCGCTCTCGTCGCTCGCGCCCTCGCAGTG 52
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db   464  CTGTTGGCTGCGCTCTCAACCCCTTGTGAATCGCGCCCGCATTTG 512

RESULT 17
AF232003      2898 bp  DNA  linear  BCT 23-MAY-2000
LOCUS      Pseudomonas syringae pv. syringae strain 61 trna-Leu gene, complete
DEFINITION  sequence; putative type III chaperone and type III effector protein
            (hrma) genes, complete cds; and HrpK (hrpK) gene, partial cds.
ACCESSION  AF232003
VERSION    AF232003.1 GI:8037767
KEYWORDS   Pseudomonas syringae pv. syringae
SOURCE     Pseudomonas syringae pv. syringae
ORGANISM   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
REFERENCE   1 (bases 1 to 2898)
AUTHORS     Alfano,J.R., Klm,H.S., Delaney,T.P. and Collmer,A.
TITLE       Evidence that the Pseudomonas syringae pv. syringae hrp-linked hrma
            gene encodes an Avr-like protein that acts in an hrp-dependent
            manner within tobacco cells
JOURNAL     Mol. Plant Microbe Interact. 10 (5), 580-588 (1997)
MEDLINE     97348579
PUBMED     9204563
REFERENCE   2 (bases 1 to 2898)
AUTHORS     Alfano,J.R., Charkowski,A.O., Deng,W.L., Badel,J.L.,
            Petnicki-Ocwieja,T., van Dijk,K. and Collmer,A.
TITLE       The Pseudomonas syringae Hrp pathogenicity island has a tripartite
            mosaic structure composed of a cluster of type III secretion genes
            bounded by exchangeable effector and conserved effector loci that
            contribute to parasitic fitness and pathogenicity in plants
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4856-4861 (2000)
MEDLINE     20243785
PUBMED     10781092
REFERENCE   3 (bases 1 to 2898)
AUTHORS     Alfano,J.R. and Collmer,A.
TITLE       Direct Submission
JOURNAL     Submitted (07-FEB-2000) Dept. Biol. Sci., UNLV, 1854 Maryland
            Parkway, Las Vegas, NV 89154, USA
FEATURES    Location/Qualifiers
             1. .2898
                /organism="Pseudomonas syringae pv. syringae"
                /mol_type="genomic DNA"
                /strain="61"
                /db_xref="taxon:321"
                /note="pathovar: syringae"
                /product="trna-Leu"
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /product="putative type III chaperone"
                /protein_id="AAF71480.1"
                /db_xref="GI:8037768"
                /translations="MEMPALAFDDKGCACNMIDKAFALTLLRDDTHORLLLTGLLEP
                HEDLPLORLLAGALNPVNVNGPGIGHDEQSLGYHAYQSIPREKVSVMLEKLEIAGLVE
                WMKCWEART"
                /note="ORF1"
                /db_xref="taxon:321"
                /note="pathovar: syringae"
                /product="trna-Leu"
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /product="putative type III effector protein"
                /protein_id="AAF71481.1"
                /db_xref="GI:8037769"
                /translations="MNPPIHARFSSVEALRHSNVDIQAIKSEGOLEVNGKRYETRAAAD
                GSIAVLRPDQSKADKFFKGAARLIGGSQRAQIAQVINEKAAAVPRLDMLGRFDL
                EKGSSAVGAIAKAADSRLLTSKQTAFSQQWAEKAEALGRYRNRYLHDLQEGHARHNA
                YECGRVKNITWKYRLSITRKLTSLYAPQIHDDREBELDLGRYIAEDRNARTGFRMV

```

	gene	PKDQAPETNSGRLTTTCVPEKYCAOLALAMATLMDKHKSVTCKVVGPYAKYCGQTDSH ILYDGLAAVKLGKLRKLSSIGPPGEFVEHTPLSMQSTGTCLSLTAESVEGGPSSHSG QARTWHIMADLKOGPMERLNKKWALAERGYPDENPALARN" complement(2849..>2898) /gene="hrpK"
CDS		complement(2849..>2898) /genes="hrpK" /codon_start=3 /transl_table=11 /product="HrpK" /protein_id="AAF71482.1" /db_xref="GI:8037770" /translation="NVDPVLDQFGIDRAH"
BASE COUNT	768 a	759 c 749 g 622 t
ORIGIN		
	Query Match	48.1%; Score 25; DB 1; Length 2898;
	Best Local Similarity	69.4%; Pred.No. 2.2e+03;
Matches	34; Conservative	0; Mismatches 15; Indels 0; Gaps 0;
Oy	4	CAGTCTCTCACGCTCATCGTGCTCTCGTCGCCTCCGCCACTGCCAGTG 52
Dd	864	CTGTTCCTGGCCTCTCAACCCCCTCTGAATTGCCGCCCCCGCATTTG 912
RESULT 38		
AF461561/c		3076 bp DNA linear BCT 30-APR-2003
LOCUS		
DEFINITION	Pseudomonas syringae pv. syringae exchangeable effector locus,	
	partial sequence.	
ACCESSION	AF461561	
VERSION	AF461561.1 GI:30231120	
KEYWORDS		
SOURCE	Pseudomonas syringae pv. syringae	
ORGANISM	Pseudomonas syringae pv. syringae	
	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;	
	Pseudomonadaceae; Pseudomonas.	
REFERENCE	1 (bases 1 to 3076)	
AUTHORS	Deng,W.L., Rehm,A.H., Charkowski,A.O., Rojas,C.M. and Collmer,A.	
TITLE	Pseudomonas syringae Exchangeable Effector Loci: Sequence Diversity	
	In Representative Pathovars and Virulence Function in P. syringae	
JOURNAL	pv. syringae B728a	
MEDLINE	J. Bacteriol. 185 (8), 2592-2602 (2003)	
PUBMED	22557902	
REFERENCE	2 (bases 1 to 3076)	
AUTHORS	Deng,W.L., Rehm,A.H., Charkowski,A.O. and Collmer,A.	
TITLE	Direct Submission	
JOURNAL	Submitted (21-DEC-2001) Plant Pathology, Cornell University, Plant	
	Science Building Rm 334, Ithaca, NY 14853, USA	
FEATURES	Location/Qualifiers	
source	1..3076	/organism="Pseudomonas syringae pv. syringae"
		/mol_type="genomic DNA"
		/strain="226"
		/db_xref="taxon:321"
		/notes"pathovar: syringae"
gene	<1..276	/gene="hrpK"
CDS	<1..276	/genes="hrpK"
		/codon_start=1
		/transl_table=11
		/product="HrpK"
		/protein_id="AAP23132.1"
		/db_xref="GI:30231121"
		/translations:"MGREASAGLGRGGQMIGRAVLAGEATGLAAGAALGAIAIPVI GWADGMALGFGISAIIDA VKHKHA QKA PDHNVP VLDQFGIDRAH"
		complement(892..2019)
CDS	/note="orf2; similar to Pseudomonas syringae pv. syringae	
	61 HoppPsyA"	
	/codon_start=1	
	/transl_table=11	

REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

Marburg, FRG
Revised by [3]
3 (bases 1 to 3719)
Vorholt, J.A.
Direct Submission
Submitted (15-JAN-1997) J.A. Vorholt, Max-Planck-Institut f.
terrestrische Mikrobiologie, Karl-von-Frisch-Strasse, D- 35043
Marburg, FRG
On Jan 16, 1997 this sequence version replaced gi:1419620.
Location/Qualifiers
1. .3719
/organism="Methanopyrus kandleri"
/mol_type="genomic DNA"
/strain="AV19/DSM 6324"
/db_xref="taxon:2320"
/clone_lib="lambda ZAP"
185. .940
/gene="orf6"
185. .940
/gene="orf6"
/codon_start=1
/transl_table=11
/protein_id="CAA67409.1"
/db_xref="GI:1781025"
/db_xref="SPTREMBL:P94940"
/translations="MRVLGVGDSGSHLKCAIVEDGSLEDHTVVESTGPKVKVLRRLALD
ELGAGIDFEDVTAVTCYGREALSDFETVPPELPAVALGASQLVEGARTVIDVGGQDT
KMKVEDGKVVDFQVNDKCAAGTRFVENVCRRLGIEMSEVDEHAGADDPVKINSMC
AVFAETEVLISLVNRGIDVERILLGLVDSVAERVATMDIKVSPEPEVVLVGMARCRVF
AELSLDRLEMGINVPNEAHVAGAFGAALWVLEK"
1043. .2347
/gene="fwbC"
1043. .2347
/gene="fwbC"
/EC_number="1.2.99.5"
/codon_start=1
/transl_table=11
/product="beta-subunit of formylmethanofuran
dehydrogenase"
/protein_id="CAA67410.1"
/db_xref="GI:1781026"
/db_xref="SPTREMBL:P94941"
/translations="MGRKVIKQVVCFCGTLCDLDDLEVVVEDGEIVVEVHACRIGAAKFP
LTAQEDHRRTEPMKNGEKKIDYEDAAETARILVEAKWPLMYGWSSTLCSEAHEVG
IEIAKGVADINTASVCHGPGSTGLQDAGVPCSTILGEVKNRADVIIYMGCPNMAHP
RHSRYTAFTRGREDRTIIIVDPRTATABELADVHLQWRPNEDVELISALRA
AHVGIETEEEVAGIPEVAEVEDLIKEASFGLTFWGGITWMSGRHRNIDNAICLI
RDLNEVAKWTLIMRGHYNVTGFNEVLTTWTTGYPFADFSRGYPVRYNCGEFSADVLI
RGEVDAFVLIASDPCGHPFRKAVEHMAWIPLVCDPHWPTTAEIADLYVFTIAGIEW
EGTAYRMSVPIINRKKVPEPPESMLNDVFELEWVIEKVBEMV"
complement(2355. .3620)
/gene="orf7"
complement(2355. .3620)
/gene="orf7"
/codon_start=1
/transl_table=11
/protein_id="CAA67411.1"
/db_xref="GI:1781027"
/db_xref="SPTREMBL:P94942"
/translations="MDIRIWKELGIDLERHDELLEALPEVVEEIIFLSOENRPERMSY
FOEVADIHGARVELRYEMROEKPILSTFCVYVPEIIVLAGGVCGVGLCGAEFFIP
DAEKLPRDLCLIKSFGFLVSRLCPCYQVATVIGENTCDGKKQKYEIMSEHKDVI
VMEPMQVDEGLEVYHEQLIKPEFVEELSNEITYESLLDAIERVNAKREAFRLY
ELRKHDPAPISDQDLNLAQIAIPYDDVDRFTKEVNELEKREVEEGEVAEAPRI
LVAGTQPMPIPHWKLKLVYSCGAVVCEESCTCTRYFEREVSTEGDDVDLKNIAEA
YMETKCAIFTPNDVRKDIKKYKWNKCDGVILYLNKFCQPTAVESHKIESLRREGI
PALKLESDSBEDVEQLKTRITFLESIA"
BASE COUNT 808 a 996 c 1060 g 855 t
ORIGIN

Query Match 48.1%; Score 25; DB 1; Length 3719;
Best Local Similarity 69.4%; Pred. No. 2.le+03;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy

Db

RESULT 40

AE010324

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

gene

CDS

gene

CDS

1 ATGCAGTTCTCTCACGCTCTCATCGCTCTCGTCTCGTCCGCGCTCCGCCA 49

418 AGGAAGCTCGGCACGGCTCTCATCGAACTCGTCTCGCTCAGACCTCCCTA 370

AE010324 10264 bp DNA linear BCT 01-AUG-2002
Methanopyrus kandleri AV19 section 23 of 157 of the complete genome

AE010324 AE009439
AE010324.1 GI:19886587

Methanopyrus kandleri AV19
Methanopyrus kandleri AV19

Archaea; Euryarchaeota; Methanopyri; Methanopyrales;
Methanopyraceae; Methanopyrus.

Slesarev,A.I., Mezhevaya,K.V., Makarova,K.S., Polushin,N.N.,
Shcherbinina,O.V., Shakhova,V.V., Belova,G.I., Aravind,L.,
Natale,D.A., Rogozin,I.B., Tatusov,R.L., Wolf,Y.I., Stetter,K.O.,
Malykh,A.G., Koonin,E.V. and Kozyavkin,S.A.
The Complete Genome of the Hyperthermophile Methanopyrus kandleri
AV19 and Monophyly of Archaeal Methanogens
Unpublished

2 (bases 1 to 10264)

Slesarev,A.I., Mezhevaya,K.V., Makarova,K.S., Polushin,N.N.,
Shcherbinina,O.V., Shakhova,V.V., Belova,G.I., Aravind,L.,
Natale,D.A., Rogozin,I.B., Tatusov,R.L., Wolf,Y.I., Stetter,K.O.,
Malykh,A.G., Koonin,E.V. and Kozyavkin,S.A.
Direct Submission
Submitted (04-FEB-2002) Fidelity Systems, Inc., Gaithersburg, MD
20879

Location/Qualifiers

1. .10264
/organism="Methanopyrus kandleri AV19"
/mol_type="genomic DNA"
/strain="AV19"
/db_xref="taxon:190192"
complement(82. .837)
/gene="MK0260"
complement(82. .837)
/gene="MK0260"
/codon_start=1
/transl_table=11
/product="Activator of 2-hydroxyglutaryl-CoA dehydratase,
contains a HSP70-class ATPase domain"
/protein_id="AAM01477.1"
/db_xref="GI:19886588"
/translation="MRVLGVGDSGSHLKCAIVEDGSLEDHTVVESTGPKVKVLRRLALD
ELGAGIDFEDVTAVTCYGREALSDFETVPPELPAVALGASQLVEGARTVIDVGGQDT
KMKVEDGKVVDFQVNDKCAAGTRFVENVCRRLGIEMSEVDEHAGADDPVKINSMC
AVFAETEVLISLVNRGIDVERILLGLVDSVAERVATMDIKVSPEPEVVLVGMARCRVF
AELSLDRLEMGINVPNEAHVAGAFGAALWVLEK"
1028. .1489
/gene="MK0261"
1028. .1489
/gene="MK0261"
/codon_start=1
/transl_table=11
/product="Uncharacterized membrane protein, conserved in
archaea"
/protein_id="AAM01478.1"
/db_xref="GI:19886589"
/translation="MARVSPPLTVGLVAGVSAALIQAVFKVPPPPAYGICIACTHRLIV
NMIVNHAAGTTLGMAPVSPFPVLTGVGIFIGALIAFAHGFRIIRTRTHPIGLFVLG
FLVINFAFMGGCPRTSLRTSLRTMGVDVALLGLISIAVGILGAFFYLKRKA"
1504. .1977
/gene="MK0262"
1504. .1977
/gene="MK0262"
/codon_start=1

```

/transl_table=11
/product="Predicted transporter component"
/protein_id="AA001479.1"
/db_xref="GI:19886590"
/translations="MAEYLHAITGLAFGLVIGLIGLQRSAMCFICGIRGVVLLRDTWL
V
OGLIGLFLVFGAGMFKFPWFLYKGSAPGDLVKGKPGFAHAATVIG
GLGVGLSVVGGCPFRNYVMAAEGNVTAAYLLGVGVGFHACTIVPIFGPK"
1982. 2560
/gene="MK0263"
CDS
1982. 2560
/gene="MK0263"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein conserved in archaea"
/protein_id="AA001480.1"
/db_xref="GI:19886591"
/translations="MEMGLSLRSLRSFSDRGPSDEREGKGLIVFESARDAMKAERTL
KEAGYDVRVAPPEIREGCDLAIYDLDVDEGVVRRTLEBIGVEPLKFVSLDEPSLKP
IELVRVKEVGYIMVRCKNMKVTVDRTGTIVNVSGGCPDPVPYLAHELVGKNVLELSE
NSMPASLGYTLCAYTLDKAVRKAKEVLLGSGS"
2680. 3381
/gene="MK0264"
CDS
2680. 3381
/gene="MK0264"
/codon_start=1
/transl_table=11
/product="Predicted sugar kinase"
/protein_id="AA001481.1"
/db_xref="GI:19886592"
/translations="MAAARVLVEFLGDEDEPRLLITAGDLGEGDGSLEIYRRLEVEDDD
VLVHIKPKIDIRRVDTSGKVIADAGMYAAKAGIGPEFHLFLPDPGLAFLADE
KAHPAYVRGFIAEVEDEEVPVSLRAYBEDVDVYVWVKGRDHVVRHEVETVDE
PLVAMECIGGTGDTLTGIVAAIIAAGFTGEACVIGCRVNRRLGEIANNATPTTRIE
LTRAIFALCEBURE"
complement(3387. 5357)
/gene="hdrA_2"
/notes="synonym: MK0265"
complement(3387. 5357)
/gene="hdrA_2"
/codon_start=1
/transl_table=11
/product="Heterodisulfide reductase, subunit A,
polyferredoxin"
/protein_id="AA001482.1"
/db_xref="GI:19886593"
/translations="MAEEPRIGYVCHVCVNTAGVVDVKEVAEFAKTLKNVVVARDY
KYVCDPQBIIQRDIEKDLNVRVAAACPRUHEPTFRRCVEEAGLNPYCFEMANIR
EHCWSWMDPPARATEKADOLVRMAKARLLESLETIKYDVTDRALVIGGVSGIQA
ALDADMGEFVILVEKEPSIGGRMAQDKTFTPDGSCICILAPKMDVSKHPNIMWT
YAEVVEVDGVGNFTVKEIKKPRYVDEDACTGCGACAEVCI EVPNFEDEGLGMRKAI
YKPFPOAIVRSVFTIDEEHCIRCGLCVEVCADAI DFOEPEI VEEVEGATICAIGYDT
CDTEREEYGVYDNVITSIELERLINASGPTGGKVVPSDCKKPKRTAFIOCVSR
DPHRTNPGVNCVMAKUAQILREKYPETQIDITMDVRAFSGKYEYERSKOY
GIRFIRGRPAEIVDEPKNLIVRAEDTLGVDVREYDLVLSVGVNPRDSADVIOE
VLSIRSPDGFMEAPKLPVPVTAIDIGIFLAGACQPKDIPSSVAQSGAAAAAATA
LAAGEVAVEPVEVEDEEICGGGTCVCLCPYGAIELVEKDGKLVAEVTAALCKGCGT
CAAAFCPGAMEQNHFKTEOLYKOIEGAFRPA"
complement(5364. 6812)
/gene="MK0266"
complement(5364. 6812)
/gene="MK0266"
/codon_start=1
/transl_table=11
/product="Coenzyme F420-reducing hydrogenase, alpha
subunit"
/protein_id="AA001483.1"
/db_xref="GI:19886594"
/translations="MRGVGTLSMVNVNVEPVRVEGHGKIVVEFDEGDLKDFHW
EVRFKFLGRIEDAPILTPRICGVVAHLSAKAADEVEGVEPEPAPKRLREL
MHAATVSHALHFVLAAPDLVKEEDPMKHEVLALARENVEVVKCAIELKIGQT
IIVAVGKPIHPVTAIVPGVSKFEDREELLELARSVELAERTVDLAELTRLE
EGGLELSELYSHMGVSDVGHLYDGVIRVVDPGNGVDREPEFEPSELDHIAEAVRP
YSYLKFPYLRDKGEEBGLRVNLTSLRNVCDRMTAPRAQRYBELVDEYKPKCHHPML

```

```

YHYARTIELSSAERCVELLEDEITGDDVREEVDPPDVTGEGIGCVCEAPRGTLIIHF
KTEBGLITEVNLIVATQNNPAMDGLGVKVAEEYLRSPEDASPEVLNRMEMVIRAYD
PCLSCATHVLGERPRITLTVEHVRRLVRIVEG"
complement(6809. 7723)
/gene="MK0267"
complement(6809. 7723)
/gene="MK0267"
/codon_start=1
/transl_table=11
/product="Coenzyme F420-reducing hydrogenase, gamma
subunit"
/protein_id="AA001484.1"
/db_xref="GI:19886595"
/translations="MGVRLVKIATITWLCSCGCHSVLDLHERLLDLLGDAELVHC
PVLMDTKPEIPEDVDVLLIEGGIRNEENVEAEFRERAEIVAVGTCACTGGVPGLAN
LVNELLRTVYVETASTENEDGIPSEDVPELTWRVPLSDVIDVYELPGCPPEPD
LIADATAILDGRPELSTNLCECPKRKEETVIRIRRVGEPEPDRLCLLQGGVP
CMGPAITRAGGACARCPAGVPCACGAPAGEIPDQGAEMSAIASIPRADVDVDPSEL
VESVPDVGMFVRFTLAGSLMPFRVDE"
complement(7720. 8145)
/gene="flpD_1"
/notes="synonym: MK0268"
complement(7720. 8145)
/gene="flpD_1"
/codon_start=1
/transl_table=11
/product="Coenzyme F420-reducing hydrogenase, delta
subunit"
/protein_id="AA001485.1"
/db_xref="GI:19886596"
/translations="MGEMPKIIFACNMCNSYGGADTAGVGRMOPSTRIIRVMCSG
RIEPSLVKAFREGADGVFVGGCHLDGCHYQSNYKWRRAELLKKFLAEIGIEPERF
RWIWSASGSEKFAEYVTEFHQTLLLEPLLELRKRL"
8285. 9712
/gene="MK0269"
8285. 9712
/gene="MK0269"
/codon_start=1
/transl_table=11
/product="Predicted membrane protein"
/protein_id="AA001486.1"
/db_xref="GI:19886597"
/translations="MKALQSLSSVAENRAGNWDKLAGIRVGRGRTVTFRRHILLA
SLACLTLPATCNVEEFESKVESLIONPFOGTKEITLVASGYSYMTGRIKVSYRGP
EEVEVEVKLNGKLKELDEDPLAGNEPIKDLGALKKFDSDVLEKLEKKTQGL
LKAACEIDVDMVYGETPTPSPDLQILQQLRTEGELHLSYKIPYPAHYVYVSVE
LLDRANLPIRAVFSYDYPNDGSKVSVDTHEKFTIETTDNIYNLIVVELTLHGGL
QVEKLAELSSESSEDLQKLENTVEKVECYPRGRTAEINGLPLMVDEGAAIIVYAV
GSPNTINSIVASKIAELMGKPLVTLGNADLOVGTGKTLVIKGSAPGGLVVIYGPAG
PONTLDVTSRGTQLVARLAGVQVPLVSLFGVILGPLLOSITTTTYPAPSGELRMSF
RGIGMATPAVIAALLGVHGRRSK"
complement(9674. 10114)
/gene="MK0270"
complement(9674. 10114)
/gene="MK0270"
/codon_start=1
/transl_table=11
/product="Uncharacterized conserved protein"
/protein_id="AA001487.1"
/db_xref="GI:19886598"
/translations="MILKFDHWSGPLYLHCSIRRGDPRIIRSSSTGVIQALSTEVA
DYRVIASACRAIRAFESDRNLARSIDLFLVRLTGTROIREAIDRAEPGDFVLVVA

```

```

Query Match      48.1%; Score 25; DB 1; Length 10264;
Best Local Similarity 69.4%; Pred. No. 1.9e+03;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 ATGCAGTCTCTCACGCTCTCATCGCTCTCGTCCGTCGCGCCGCTCGCA 49
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 604 AGGAAGCTCGGCACGGTCTCTCAAGAACTCGTCGCTCGCTCCTCCTA 652
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: November 13, 2003, 08:17:42

```

Job time : 1954 secs
